GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2000, 22:49:52 ; Search time 1394.2 Seconds
(without alignments)
-4695.484 Million cell updates/sec

Title: US-09-304-121-1
Perfect score: 2156
Sequence: 1 cgggcccgttgcaagatggc.....aaaaaaaaaaaaaaaaaaaa 2156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2156	100.0	2156	5	AR009723	AR009723 Sequence
2	2156	100.0	2156	9	HUMHSF1	M64673 Human heat
3	1218.2	56.5	1947	12	MMHSF1	X61753 M.musculus
4	1079.6	50.1	1647	12	RRHSF1	X83094 R.rattus mR
5	808	37.5	1774	4	CHKHSF3A	L06098 Chicken mRN
c 6	599	27.8	131973	44	AF205589	AF205589 Homo sapi
7	476	22.1	1657	4	XELHSF	L36924 Xenopus lae
8	277.4	12.9	1555	9	D87673	D87673 Homo sapien
9	277.4	12.9	1577	10	AB029348	AB029348 Homo sapi
10	259.6	12.0	1591	12	AB029350	AB029350 Mus muscu
11	259.6	12.0	1686	12	AB029349	AB029349 Mus muscu
12	241	11.2	11395	12	AF059275	AF059275 Mus muscu
13	230.4	10.7	2675	4	CHKHSF3C	L06126 Chicken mRN
14	204.2	9.5	1972	12	MMHSF2	X61754 M.musculus
15	203.6	9.4	2366	4	CHKHSF3B	L06125 Chicken mRN
16	199.4	9.2	1792	12	AF172640	AF172640 Rattus no
17	197.4	9.2	2411	9	HUMHSF2	M65217 Human heat
18	183.4	8.5	2760	34	DROHSPHEX	M60070 D.melanogas
19	183.4	8.5	2781	5	AR009733	AR009733 Sequence
20	180.6	8.4	7597	12	AF061503	AF061503 Mus muscu
21	103.2	4.8	61267	35	AC004336	AC004336 Drosophil
22	100.8	4.7	2048	7	ZMHSFB	X82943 Z.mays mRNA
23	92	4.3	1560	35	AF043416	AF043416 Schistoso
24	92	4.3	2369	35	AF043418	AF043418 Schistoso
25	89	4.1	1410	7	GMHSF34	Z46953 G.max mRNA
26	88.8	4.1	2261	35	AF043417	AF043417 Schistoso
27	86.6	4.0	3340	7	KLHSFG	X55149 K. lactis H
28	86	4.0	5221	10	AB029347	AB029347 Homo sapi
29	84.2	3.9	1625	7	AB014483	AB014483 Nicotiana
30	82.4	3.8	1690	7	AB014484	AB014484 Nicotiana
31	78	3.6	1203	7	GMHSF5	Z46956 G.max mRNA
32	76.8	3.6	1791	7	ATHSF3	Y14068 Arabidopsis
33	76.8	3.6	2251	7	YSPHSF	M94683 Schizosacch
34	76.2	3.5	1530	7	LPHSF24	X55347 Tomato mRNA
35	75.2	3.5	3319	7	ATHSF1	X76167 A.thaliana
c 36	75.2	3.5	81835	8	ATFCA9	Z97344 Arabidopsis
37	74.2	3.4	1791	7	LPHSF8	X67600 L.peruvianu
38	73.6	3.4	671	7	GMHSF21	Z46952 G.max mRNA
39	69.4	3.2	1407	8	ATU58561	U68561 Arabidopsis
40	67.8	3.1	1315	7	LPHSF30	X67601 L.peruvianu
41	67.4	3.1	9005	7	LEHSF8	X67599 L.esculentu
42	67.2	3.1	39868	45	AC016887	AC016887 Homo sapi
c 43	66.6	3.1	167945	44	AC013364	AC013364 Homo sapi
44	66.4	3.1	31827	7	AB018115	AB018115 Arabidops
c 45	66.4	3.1	139207	41	AC011429	AC011429 Homo sapi

ALIGNMENTS

RESULT 1
AR009723
LOCUS AR009723 2156 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 31 from patent US 5756343.
ACCESSION AR009723
VERSION AR009723.1 GI:3968528
KEYWORDS .

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2156)
  AUTHORS   Wu,C., Clos,J., Westwood,J.Timothy and Rabindran,S.
  TITLE     Cell stress transcriptional factors
  JOURNAL   Patent: US 5756343-A 31 26-MAY-1998;
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="unknown"
BASE COUNT      435 a    739 c    628 g    354 t
ORIGIN
```

Query Match 100.0%; Score 2156; DB 5; Length 2156;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 cgggcccgttgcaagatggcggcggccatgctgggccccggggctgtgtgtgcgcagcgg 60
Db       1 CGGGCCCGTTGCAAGATGGCGGCGGCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGG 60

Qy      61 gcggcggcgcggcccggaaggctggcgcggcgacggcgttagcccggccctcggcccctc 120
Db      61 GCGGCGGCGCGGCCCGGAAGGCTGGCGCGGCGACGGCGTTAGCCCGGCCCTCGGCCCCTC 120

Qy     121 tttgcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccc 180
Db     121 TTTGCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCC 180

Qy     181 cggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgag 240
Db     181 CGGCGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAG 240

Qy     241 cgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgtt 300
Db     241 CGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTT 300

Qy     301 cgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggc 360
Db     301 CGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

Qy     361 cagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcaggg 420
Db     361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGG 420

Qy     421 cggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtgg 480
Db     421 CGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGG 480

Qy     481 ccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagag 540
Db     481 CCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAG 540

Qy     541 tgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgat 600
Db     541 TGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGAT 600

Qy     601 gaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggc 660
Db     601 GAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

Qy     661 tctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaa 720
Db     661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAA 720

Qy     721 caagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagag 780
Db     721 CAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGAG 780

Qy     781 aaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccg 840
Db     781 AAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCCG 840

Qy     841 gcagttctccctggagcacgtccacggctcgggcccctactcggcccctccccagcctа 900
Db     841 GCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTA 900

Qy     901 cagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctccga 960
Db     901 CAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

Qy     961 catcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacgagag 1020
Db     961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCATAGACGAGAG 1020

Qy    1021 gcccctatccagcagcccccctggtgcgtgtcaaggaggagcccccccagcccgcctcagag 1080
Db    1021 GCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCCCCCAGCCCGCCTCAGAG 1080

Qy    1081 cccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtcccc 1140
Db    1081 CCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCCATCTTCCGTGGACACCCTCTTGTCCCC 1140

Qy    1141 gaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagc 1200
Db    1141 GACCGCCCTCATTGACTCCATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGC 1200

Qy    1201 cctcacggacgccaggggccacacggacaccgagggccggcctccctcccccccgcccac 1260
Db    1201 CCTCACGGACGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

Qy    1261 ctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgacca 1320
Db    1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGCTCAGTGACCA 1320

Qy    1321 cttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacggctt 1380
Db    1321 CTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACCATGCTGAGCAGCCACGGCTT 1380

Qy    1381 cagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacat 1440
Db    1381 CAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACAT 1440

Qy    1441 gagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctccccagga 1500
Db    1441 GAGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGA 1500

Qy    1501 gccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgca 1560
Db    1501 GCCCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

Qy    1561 ctacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacga 1620
Db    1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGA 1620

Qy    1621 cctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcgc 1680
Db    1621 CCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCGC 1680

Qy    1681 cgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccccac 1740
Db    1681 CGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCAC 1740

Qy    1741 tgtctcctagaggccccggaggagctgggccagccgcccaccccacccccagtgcaggg 1800
Db    1741 TGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGG 1800

Qy    1801 ctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgccata 1860
Db    1801 CTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATA 1860

Qy    1861 gccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcaccct 1920
Db    1861 GCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCCT 1920
```

```
Qy   1921 ggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttcacagc 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAGC 1980

Qy   1981 cacacctggactgaccctgcaggttgttcatagtcagaattgtattttggatttttacac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACAC 2040

Qy   2041 aactgtcccgttccccgctccacagagatacacagatatatacacacagtggatggacgg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACGG 2100

Qy   2101 acaagacaggcagagatctataaacagacaggctctaaaaaaaaaaaaaaaaaaaaa 2156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAAA 2156


RESULT   2
HUMHSF1
LOCUS       HUMHSF1      2156 bp    mRNA            PRI       08-NOV-1994
DEFINITION  Human heat shock factor 1 (TCF5) mRNA, complete cds.
ACCESSION   M64673
VERSION     M64673.1  GI:184402
KEYWORDS    DNA-binding transcription factor; heat shock factor 1.
SOURCE      Homo sapiens lymphoid cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2156)
  AUTHORS   Rabindran,S.K., Giorgi,G., Clos,J. and Wu,C.
  TITLE     Molecular cloning and expression of a human heat shock factor, HSF1
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (16), 6906-6910 (1991)
  MEDLINE   91334376
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="B cell"
                     /tissue_type="lymphoid"
                     /map="20"
     gene            161. .1750
                     /gene="TCF5"
     CDS             161. .1750
                     /gene="TCF5"
                     /codon_start=1
                     /db_xref="GDB:G00-126-373"
                     /product="heat shock factor 1"
                     /protein_id="AAA52695.1"
                     /db_xref="GI:184403"
                     /translation="MDLPVGPGAAGPSNVPAFLTKLWTLVSDPDTDALICWSPSGNSF
                     HVFDQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVHIEQGGLVKPERDDTEFQHP
                     CFLRGQEQLLENIKRKVTSVSTLKSEDIKIRQDSVTKLLTDVQLMKGKQECMDSKLLA
                     MKHENEALWREVASLRQKHAQQQKVVNKLIQFLISLVQSNRILGVKRKIPLMLNDSGS
                     AHSMPKYSRQFSLEHVHGSGPYSAPSPAYSSSSLYAPDAVASSGPIISDITELAPASP
                     MASPGGSIDERPLSSSPLVRVKEEPPSPPQSPRVEEASPGRPSSVDTLLSPTALIDSI
                     LRESEPAPASVTALTDARGHTDTEGRPPSPPPTSTPEKCLSVACLDKNELSDHLDAMD
                     SNLDNLQTMLSSHGFSVDTSALLDLFSPSVTVPDMSLPDLDSSLASIQELLSPQEPPR
                     PPEAENSSPDSGKQLVHYTAQPLFLLDPGSVDTGSNDLPVLFELGEGSYFSEGDGFAE
                     DPTISLLTGSEPPKAKDPTVS"
BASE COUNT      435 a    739 c    628 g    354 t
ORIGIN


  Query Match             100.0%;  Score 2156;  DB 9;  Length 2156;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgggcccgttgcaagatggcggcggccatgctgggccccggggctgtgtgtgcgcagcgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGGCCCGTTGCAAGATGGCGGCGGCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGG 60

Qy     61 gcggcggcgcggcccggaaggctggcgcggcgacggcgttagccccggccctcggccccctc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     61 GCGGCGGCGCGGCCCGGAAGGCTGGCGCGGCGACGGCGTTAGCCCGGCCCTCGGCCCCTC 120

Qy    121 tttgcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCC 180

Qy    181 cggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGGCGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAG 240

Qy    241 cgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgtt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTT 300

Qy    301 cgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

Qy    361 cagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcaggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGG 420

Qy    421 cggcctggtcaagccagagagagagacgacacggagttccagcacccatgcttcctgcgtgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGG 480

Qy    481 ccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagag 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAG 540

Qy    541 tgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGAT 600

Qy    601 gaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

Qy    661 tctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAA 720

Qy    721 caagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGAG 780

Qy    781 aaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCCG 840

Qy    841 gcagttctccctggagcacgtccacggctcgggcccctactcggccccctccccagccta 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTA 900

Qy    901 cagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctccga 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

Qy    961 catcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacgagag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCATAGACGAGAG 1020

Qy   1021 gcccctatccagcagccccctggtgcgtgtcaaggaggagcccccccagcccgcctcagag 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCCCCCAGCCCGCCTCAGAG 1080

Qy   1081 cccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtcccc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCCATCTTCCGTGGACACCCTCTTGTCCCC 1140

Qy   1141 gaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GACCGCCCTCATTGACTCCATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGC 1200
```

```
Qy   1201 cctcacggacgccaggggccacacggacaccgagggccggcctccctccccccgcccac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTCACGGACGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

Qy   1261 ctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgacca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGCTCAGTGACCA 1320

Qy   1321 cttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacggctt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACCATGCTGAGCAGCCACGGCTT 1380

Qy   1381 cagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACAT 1440

Qy   1441 gagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctccccagga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GAGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGA 1500

Qy   1501 gccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

Qy   1561 ctacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacga 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGA 1620

Qy   1621 cctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcgc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCGC 1680

Qy   1681 cgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccccac 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCAC 1740

Qy   1741 tgtctcctagaggccccggaggagctgggccagccgcccaccccccacccccagtgcaggg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGG 1800

Qy   1801 ctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgccata 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATA 1860

Qy   1861 gccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcaccct 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCCT 1920

Qy   1921 ggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttcacagc 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAGC 1980

Qy   1981 cacacctggactgaccctgcaggttgttcatagtcagaattgtattttggatttttacac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACAC 2040

Qy   2041 aactgtcccgttccccgctccacagagatacacagatatatacacacagtggatggacgg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACGG 2100

Qy   2101 acaagacaggcagagatctataaacagacaggctctaaaaaaaaaaaaaaaaaaaaa 2156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAA 2156

RESULT   3
MMHSF1
LOCUS       MMHSF1       1947 bp    mRNA            ROD       28-NOV-1991
DEFINITION  M.musculus mRNA for heat shock transcription factor 1.
ACCESSION   X61753
VERSION     X61753.1  GI:51445
KEYWORDS    heat shock transcription factor; HSF1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1947)
  AUTHORS   Sarge,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1991) K.D. Sarge, Northwestern University, Dep of
            Biochem.,Mol. and Cell. Biology, Hogan 2-100, 2153 Sheridan Rd.,
            Evanston IL 60208, USA
REFERENCE   2  (bases 1 to 1947)
  AUTHORS   Sarge,K.D., Zimarino,V., Holm,K., Wu,C. and Morimoto,R.I.
  TITLE     Cloning and characterization of two mouse heat shock factors with
            distinct inducible and constitutive DNA-binding ability
  JOURNAL   Genes Dev. 5 (10), 1902-1911 (1991)
  MEDLINE   92009180
COMMENT     See also x61754.
FEATURES             Location/Qualifiers
     source          1. .1947
                     /organism="Mus musculus"
                     /strain="WEHI-3"
                     /db_xref="taxon:10090"
                     /dev_stage="adult"
                     /cell_type="macrophage precursor"
                     /cell_line="WEHI-3"
                     /clone_lib="lambda ZAP cDNA"
                     /clone="C12"
     CDS             140. .1651
                     /codon_start=1
                     /product="heat shock transcription factor 1"
                     /protein_id="CAA43892.1"
                     /db_xref="GI:51446"
                     /db_xref="MGD:MGI:96238"
                     /db_xref="SWISS-PROT:P38532"
                     /translation="MDLAVGPGAAGPSNVPAFLTKLWTLVSDPDTDALICWSPSGNSF
                     HVFDQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVHIEQGGLVKPERDDTEFQHP
                     CFLRGQEQLLENIKRKVTSVSTLKSEDIKIRQDSVTRLLTDVQLMKGKQECMDSKLLA
                     MKHENEALWREVASLRQKHAQQQKVVNKLIQFLISLVQSNRILGVKRKIPLMLSDSNS
                     AHSVPKYGRQYSLEHVHGPGPYSAPSPAYSSSSLYSSDAVTSSGPIISDITELAPTSP
                     LASPGRSIDERPLSSSTLVRVKQEPPSPPHSPRVLEASPGRPSSMDTPLSPTAFIDSI
                     LRESEPTPAASNTAPMDTTGAQAPALPTPSTPEKCLSVACLDKNELSDHLDAMDSNLD
                     NLQTMLTSHGFSVDTSALLDIQELLSPQEPPRPIEAENSNPDSGKQLVHYTAQPLFLL
                     DPDAVDTGSSELPVLFELGESSYFSEGDDYTDDPTISLLTGTEPHKAKDPTVS"
BASE COUNT      435 a    613 c    521 g    378 t
ORIGIN

  Query Match           56.5%;  Score 1218.2;  DB 12;  Length 1947;
  Best Local Similarity  78.9%;  Pred. No. 2e-199;
  Matches 1569;  Conservative    0;  Mismatches  323;  Indels   97;  Gaps    6;

Qy     43 gctgtgtgtgcgcagcgggcggcggcgcggcccggaaggctggcgcggcgacggcgttag 102
          |||||||||||||||||||||||| |||||||||| ||| || || ||||||| |  |||
Db     21 GCTGTGTGTGCGCAGCGGGCGGCGGAGCGGCCCGGACGGCAGGGGCAGCGACGACACTAG 80

Qy    103 cccggccctcggcccctctttgcggccgctccctccgcctattccctcctt-gctcgaga 161
          | | ||| || ||| |||||       ||  || | |||| | ||  | || |  |||||
Db     81 CTCAGCCTTCAGCCACTCTTCTAAAAGGCCACCCCAGCCTCTGCCTGCTTTCGTCCGAGA 140

Qy    162 tggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctgacca 221
          |||||||| |||||||||||| || ||||||||||||||||||||||||||||| ||||
Db    141 TGGATCTGGCCGTGGGCCCCGGTGCAGCGGGGCCCAGCAACGTCCCGGCCTTCCTAACCA 200

Qy    222 agctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccgagcg 281
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||| |
Db    201 AGCTGTGGACCCTCGTGAGCGACCCGGACACAGACGCGCTCATCTGCTGGAGCCCGAGTG 260

Qy    282 ggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgctgcccaagtact 341
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    261 GGAACAGCTTCCACGTGTTTGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACT 320

Qy    342 tcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtatggcttccggaaag 401
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||| ||||
```

```
Db    321 TCAAGCACAACAACATGGCTAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGAAAAG 380

Qy    402 tggtccacatcgagcagggcggcctggtcaagccagagagagacgacacggagttccagc 461
Db    381 TAGTCCACATTGAGCAGGGTGGCCTGGTCAAGCCTGAGAGAGATGACACCGAGTTCCAGC 440

Qy    462 acccatgcttcctgcgtggccaggagcagctccttgagaacatcaagaggaaagtgacca 521
Db    441 ATCCTTGTTTCTTGCGTGGACAGGAACAGCTCCTTGAGAACATCAAGAGGAAAGTGACCA 500

Qy    522 gtgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgtcaccaagctgc 581
Db    501 GCGTGTCCACCCTGAAGAGTGAGGACATAAAAATACGCCAGGACAGTGTCACCCGGCTGT 560

Qy    582 tgacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaagctcctggcca 641
Db    561 TGACAGATGTGCAGCTGATGAAGGGGAAACAGGAGTGTATGGACTCCAAGCTCCTGGCCA 620

Qy    642 tgaagcatgagaatgaggctctgtggcgggaggtggccagccttcggcagaagcatgccc 701
Db    621 TGAAGCACGAGAACGAGGCCCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCC 680

Qy    702 agcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggtgcagtcaaacc 761
Db    681 AGCAGCAAAAAGTTGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCGAACC 740

Qy    762 ggatcctggggggtgaagagaaagatccccctgatgctgaacgacagtggctcagcacatt 821
Db    741 GGATCCTGGGGGTGAAGAGAAAGATCCCTCTGATGTTGAGTGACAGCAACTCAGCACACT 800

Qy    822 ccatgcccaagtatagccggcagttctccctggagcacgtccacggctcgggcccctact 881
Db    801 CTGTGCCCAAGTATGGTCGACAGTACTCCCTGGAGCATGTCCATGGTCCTGGCCCATACT 860

Qy    882 cggcccctccccagcctacagcagctccagcctctacgcccctgatgctgtggccagct 941
Db    861 CAGCTCCATCTCCAGCCTACAGCAGCTCTAGCCTTTACTCCTCTGATGCTGTCACCAGCT 920

Qy    942 ctggacccatcatctccgacatcaccgagctggctcctgccagccccatggcctccccccg 1001
Db    921 CTGGACCCATAATCTCCGATATCACTGAGCTGGCTCCCACCAGCCCTTTGGCCTCCCCAG 980

Qy   1002 gcgggagcatagacgagaggcccctatccagcagcccccctggtgcgtgtcaaggaggagc 1061
Db    981 GCAGGAGCATAGATGAGAGGCCTCTGTCCAGCAGCACTCTGGTCCGTGTCAAGCAAGAGC 1040

Qy   1062 cccccagcccgcctcagagcccccgggtagaggaggcgagtcccgggcgcccatcttccg 1121
Db   1041 CCCCCAGCCCACCTCACAGCCCTCGGGTACTGGAGGCGAGCCCTGGGCGCCCATCCTCCA 1100

Qy   1122 tggacaccctcttgtccccgaccgccctcattgactccatcctgcgggagagtgaacctg 1181
Db   1101 TGGATACCCCTTTGTCCCCAACTGCCTTCATTGACTCCATCCTTCGAGAGAGCGAGCCTA 1160

Qy   1182 cccccgcctccgtcacagccctcacggacgccaggggccacacggacaccgagggccggc 1241
Db   1161 CCCCTGCTGCCTCAAACACAGCCCCTATGGACA------------CAACCGGAGCCCAAG 1208

Qy   1242 ctccctcccccccgcccacctccacccctgaaaagtgcctcagcgtagcctgcctggaca 1301
Db   1209 CCCCCGCACTCCCGACCCCCTCCACCCCTGAGAAGTGCCTCAGCGTAGCCTGCCTAGACA 1268

Qy   1302 agaatgagctcagtgaccacttggatgctatggactccaacctggataacctgcagacca 1361
Db   1269 AGAACGAGCTAAGTGATCACCTGGATGCCATGGACTCCAACCTGGACAACCTGCAGACCA 1328

Qy   1362 tgctgagcagccacggcttcagcgtggacaccagtgccctgctggacctgttcagcccct 1421
Db   1329 TGCTGACAAGCCACGGCTTCAGTGTGGACACCAGTGCCCTGCTGGAC------------- 1376

Qy   1422 cggtgaccgtgcccgacatgagcctgcctgaccttgacagcagcctggccagtatccaag 1481
Db   1376 -----------------------------------------------------ATTCAGG 1382

Qy   1482 agctcctgtctccccaggagccccccaggcctcccgaggcagagaacagcagcccggatt 1541
Db   1383 AGCTTCTGTCTCCACAAGAGCCTCCCAGGCCTATTGAGGCAGAGAACAGTAACCCCGACT 1442

Qy   1542 cagggaagcagctggtgcactacacagcgcagccgctgttcctgctggaccccggctccg 1601
Db   1443 CAGGAAAGCAGCTGGTGCACTACACGGCTCAGCCTCTGTTCCTGCTGGATCCTGATGCTG 1502

Qy   1602 tggacaccgggagcaacgacctgccggtgctgtttgagctgggagagggctcctacttct 1661
Db   1503 TGGACACAGGGAGCAGTGAGCTGCCTGTGCTCTTTGAGCTGGGGGAGAGCTCCTACTTCT 1562

Qy   1662 ccgaaggggacggcttcgccgaggaccccaccatctccctgctgacaggctcggagcctc 1721
Db   1563 CTGAGGGGGATGACTACACGGATGATCCCACCATCTCTCTTCTGACAGGCACTGAACCCC 1622

Qy   1722 ccaaagccaaggaccccactgtctcctagaggccccggagg----------agctgggcc 1771
Db   1623 ATAAAGCCAAGGACCCCACTGTCTCCTAGAGCTCTCAGGAGTTGTCAGGCTGGCTTGTGC 1682

Qy   1772 agccgcccacccccacccccagtgcagggctggtcttggggaggcagggcagcctcgcgg 1831
Db   1683 CTGGCCCCCAACCTATCCCTAGGACATGGCTGGTCCTAGGGAGACAAAACAGTTGGGTAG 1742

Qy   1832 tcttgggcactggtgggtcggccgccatagccccagtaggaca---aacgggctcgggtc 1888
Db   1743 TCCAGGGGACCCTAGGTCAAGCCACCACAACCCCAGTGGAGCACAGATGGAACTTGGTCC 1802

Qy   1889 tgggcagcacctctggtcaggagggtcaccctg-----gcctgccagtctgccttccccc 1943
Db   1803 TGGGCAGTACCTTGGATCAGGAGGAAGATCCTGAGGGCTGCATACCTGCTGCCTTTACCC 1862

Qy   1944 aaccccgtgtcctgtggtttggttggggcttcacagccacacctggactgaccctgcagg 2003
Db   1863 CAGCCCCAGGTCTACTCTCTGGTCACAGCTTCACAGCCACACTTGGACTGACCCTGCAGG 1922

Qy   2004 ttgttcata 2012
Db   1923 TTGTTCATA 1931
```

```
RESULT       4
RRHSF1
LOCUS       RRHSF1        1647 bp    mRNA            ROD       30-MAR-1995
DEFINITION  R.rattus mRNA for heat shock transcription factor 1.
ACCESSION   X83094
VERSION     X83094.1  GI:758270
KEYWORDS    heat shock transcription factor 1; HSF1 gene.
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 1647)
  AUTHORS   Swamynathan,S.K., Revathi,C.J. and Srinivas,U.K.
  TITLE     Cloning and characterization of rat heat shock transcription factor
            1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1647)
  AUTHORS   Srinivas,U.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1994) U.K. Srinivas, Centre for Cellular and
            Molecular Biology, Habsiguda, Hyderabad, PIN-500 007, INDIA
FEATURES             Location/Qualifiers
     source          1..1647
                     /organism="Rattus rattus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10117"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /cell_type="hepatocytes"
                     /clone_lib="Stratagene cDNA library"
     gene            1..1348
```

```
                     /gene="HSF1"
     CDS             <1. .1348
                     /gene="HSF1"
                     /codon_start=2
                     /product="heat shock transcription factor 1"
                     /protein_id="CAA58149.1"
                     /db_xref="GI:758271"
                     /db_xref="SPTREMBL:Q63717"
                     /translation="FRKVVHIEQGGLVKPERDDTEFQHPCFLRGQEQLLENIKRKVTS
                     VSTLKSEDIKIRQDSVTRLLTDVQLMKGKQECMDSKLLAMKHENEALWREVASLRQKH
                     AQQQKVVNKLIQFLISLVQSNRILGVKRKIPLMLSDSSSAHSVPKYGRQYSLEHVHGP
                     GPYSAPSPAYSSSSLYSSDAVTSSGPIISDITELAPTSPLASPGRSIDERPLSSSTLV
                     RVKEEPPSPPHSPRVLEASPGRPASMDTPLSPTAFIDSILRESEPTPAASNTAPMDTT
                     GAQAPAHPAPSTPEKCLSVACLDKNELSDHLDAMDSNLDNLQTMLTSHGFSVDTSALL
                     DLFSPSVTMPDMSLPDLDSSLASIQELLSPQEPPRPIEAENSNADSGKQLVHYTAQPL
                     FLLDPDAVDTGSSELPVLFELGESSYFSEGDDYTDDPTISLLTGTEPHKAKDPTVS"
     3'UTR           1349. .1647
BASE COUNT      384 a    512 c    425 g    326 t
ORIGIN
```

Query Match 50.1%; Score 1079.6; DB 12; Length 1647;
Best Local Similarity 80.4%; Pred. No. 1e-175;
Matches 1337; Conservative 0; Mismatches 294; Indels 31; Gaps 5;

```
Qy    391 cttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgacac 450
Db      1 CTTCCGAAAAGTAGTCCACATTGAGCAGGGTGGCCTGGTCAAGCCTGAGAGGGATGACAC 60

Qy    451 ggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaagag 510
Db     61 CGAGTTCCAGCATCCATGTTTCCTGCGTGGCCAGGAACAGCTCCTTGAGAATATCAAGAG 120

Qy    511 gaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgt 570
Db    121 GAAAGTGACCAGCGTGTCCACCCTGAAGAGTGAGGACATAAAAATACGCCAGGACAGTGT 180

Qy    571 caccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaa 630
Db    181 CACCAGGCTGTTGACAGATGTGCAGCTGATGAAGGGGAAACAGGAGTGTATGGACTCCAA 240

Qy    631 gctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagccttcggca 690
Db    241 GCTCCTGGCCATGAAGCACGAGAACGAGGCCCTGTGGCGGGAGGTGGCCAGCCTTCGGCA 300

Qy    691 gaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggt 750
Db    301 GAAGCATGCCCAGCAGCAAAAAGTTGTCAACAAGCTCATCCAATTCCTGATCTCACTGGT 360

Qy    751 gcagtcaaaccggatcctgggggtgaagagaaagatccccctgatgctgaacgacagtgg 810
Db    361 GCAGTCGAACCGGATCCTGGGGGTGAAAAGAAAGATCCCTCTGATGTTGAGTGACAGCAG 420

Qy    811 ctcagcacattccatgcccaagtatagccggcagttctccctggagcacgtccacggctc 870
Db    421 CTCAGCACACTCTGTGCCCAAGTATGGTCGACAGTACTCCCTGGAGCATGTCCATGGTCC 480

Qy    871 gggcccctactcggcccccctccccagcctacagcagctccagcctctacgcccctgatgc 930
Db    481 CGGCCCATACTCAGCTCCATCTCCAGCCTACAGCAGCTCTAGCCTTTACTCCTCTGATGC 540

Qy    931 tgtggccagctctggacccatcatctccgacatcaccgagctggctcctgccagccccat 990
Db    541 TGTCACCAGCTCTGGGCCCATAATCTCCGATATCACTGAGCTGGCTCCCACCAGCCCTTT 600

Qy    991 ggcctcccccggcgggagcatagacgagaggcccctatccagcagccccctggtgcgtgt 1050
Db    601 GGCCTCCCCAGGCAGGAGCATAGATGAGAGGCCTCTGTCCAGCAGCACCCTGGTCCGTGT 660

Qy   1051 caaggaggagccccccagcccgcctcagagcccccgggtagaggaggcgagtcccgggcg 1110
Db    661 CAAGGAAGAGCCCCCCAGCCCACCTCATAGCCCTCGGGTACTGGAGGCAAGCCCTGGGCG 720

Qy   1111 cccatcttccgtggacaccctcttgtccccgaccgccctcattgactccatcctgcggga 1170
Db    721 CCCAGCCTCCATGGATACCCCTTTGTCCCCAACTGCCTTCATTGACTCCATCCTTCGAGA 780

Qy   1171 gagtgaacctgcccccgcctccgtcacagccctcacggacgccaggggccacacggacac 1230
Db    781 GAGCGAACCTACCCCTGCTGCCTCAAACACAGCCCCTATGGACA-----------CAAC 828

Qy   1231 cgagggccggcctccctcccccccgcccacctccacccctgaaaagtgcctcagcgtagc 1290
Db    829 CGGAGCCCAAGCCCCTGCACACCCAGCCCCCTCCACCCCTGAGAAGTGCCTCAGCGTAGC 888

Qy   1291 ctgcctggacaagaatgagctcagtgaccacttggatgctatggactccaacctggataa 1350
Db    889 CTGCCTAGACAAGAACGAGCTAAGTGATCACCTGGATGCCATGGACTCCAACCTGGACAA 948

Qy   1351 cctgcagaccatgctgagcagccacggcttcagcgtggacaccagtgccctgctggacct 1410
Db    949 CCTGCAGACCATGCTGACAAGCCACGGCTTCAGTGTGGACACCAGTGCCCTGCTGGACCT 1008

Qy   1411 gttcagcccctcggtgaccgtgcccgacatgagcctgcctgaccttgacagcagcctggc 1470
Db   1009 GTTCAGCCCCTCGGTGACCATGCCCGACATGAGCCTGCCTGACCTGGACAGCAGCCTGGC 1068

Qy   1471 cagtatccaagagctcctgtctccccaggagccccccaggcctcccgaggcagagaacag 1530
Db   1069 CAGCATTCAGGAACTTCTATCTCCCCAAGAGCCTCCCAGGCCTATTGAGGCAGAGAATAG 1128

Qy   1531 cagcccggattcagggaagcagctggtgcactacacagcgcagccgctgttcctgctgga 1590
Db   1129 TAACGCCGACTCAGGAAAGCAGCTGGTGCACTACACGGCTCAGCCTCTGTTCCTGCTGGA 1188

Qy   1591 ccccggctccgtggacaccgggagcaacgacctgccggtgctgtttgagctgggagaggg 1650
Db   1189 CCCTGATGCTGTGGACACAGGGAGCAGTGAACTGCCTGTGCTCTTTGAGCTGGGGGAGAG 1248

Qy   1651 ctcctacttctccgaaggggacggcttcgccgaggaccccaccatctccctgctgacagg 1710
Db   1249 CTCCTACTTCTCTGAGGGAGATGACTACACGGATGATCCCACCATCTCTCTTCTGACGGG 1308

Qy   1711 ctcggagcctcccaaagccaaggaccccactgtctcctagaggccccggagg-------a 1763
Db   1309 CACTGAACCCCACAAAGCCAAGGACCCCACTGTCTCCTAGAGGTTTCAGTTGTCAGGCTG 1368

Qy   1764 gctgggccagccgcccacccccacccccagtgcagggctggtcttggggaggcagggcag 1823
Db   1369 GCTTGGGCCTGGCCCCCAGCCTATTCCTAGGACACGGCTGGTCCTGGGGAGACAAGACAG 1428

Qy   1824 cctcgcggtcttgggcactggtgggtcggccgccatagccccagtaggaca---aacggg 1880
Db   1429 TTGGGTAGTCCAGGGAATCCTAGGTCAAGCCAGCACAACCCCAGTGGAGCACAGATGGGA 1488

Qy   1881 ctcgggtctgggcagcacctctggtcaggagggtcaccctg-----gcctgccagtctgc 1935
Db   1489 CTTGGGCTTGGGCAGTACCTTGGATCAAGAGGAAGATCCTGAAGGCTGCATACCTGCTGC 1548

Qy   1936 cttccccaaccccgtgtcctgtggtttggttggggcttcacagccacacctggactgac 1995
Db   1549 CTTCACCCCAGCCCCAAGTCTGCTCTCTGGTCAGAGCTTCACAGCCACACTTGGACTGAC 1608

Qy   1996 cctgcaggttgttcatagtcagaattgtattttggattttta 2037
Db   1609 CCTGCAGGTTGTTCAT----AAAATTGTATTTTGATTTTTAA 1646
```

```
RESULT   5
CHKHSF3A
LOCUS       CHKHSF3A     1774 bp    mRNA            VRT       17-JUN-1993
DEFINITION  Chicken mRNA sequence.
ACCESSION   L06098
VERSION     L06098.1  GI:289814
KEYWORDS    .
SOURCE      Gallus gallus (sub_species domesticus) (library: lambda ZAPII)
```

```
          embryo blood cDNA to mRNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1774)
  AUTHORS   Nakai,A. and Morimoto,R.I.
  TITLE     Characterization of a novel chicken heat sock transcription factor,
            heat shock factor 3, suggests a new regulatory pathway
  JOURNAL   Mol. Cell. Biol. 13, 1983-1997 (1993)
  MEDLINE   93204945
FEATURES             Location/Qualifiers
     source          1. .1774
                     /organism="Gallus gallus"
                     /sub_species="domesticus"
                     /db_xref="taxon:9031"
                     /cell_type="red blood cell"
                     /dev_stage="embryo"
                     /tissue_type="blood"
                     /tissue_lib="lambda ZAPII"
BASE COUNT      443 a    582 c    479 g    270 t
ORIGIN
```

Query Match 37.5%; Score 808; DB 4; Length 1774;
Best Local Similarity 72.6%; Pred. No. 2.9e-129;
Matches 1137; Conservative 0; Mismatches 300; Indels 129; Gaps 2;

```
Qy    189 cggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgagcgacccgg 248
          ||||   |||||||||  ||||||||||||||||||||||||  | |||   ||||| |
Db     74 CGGGCGGCAGCAACGTGTCGGCCTTCCTGACCAAGCTGTGGACGTTGGTGGAGGACCCCG 133

Qy    249 acaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgttcgaccagg 308
          | || ||| |||| |||||||||||||||| || ||||||||||| |||||||||| |
Db    134 AGACGGACCCGCTGATCTGCTGGAGCCCGAGTGGAAACAGCTTCCATGTGTTCGACCAAG 193

Qy    309 gccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcg 368
          | ||||| |||||||||||||| || ||||||||||||||||||||||||||||||||
Db    194 GACAGTTCGCCAAGGAGGTGCTCCCAAAGTACTTCAAGCACAACAACATGGCCAGCTTCG 253

Qy    369 tgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagggcggcctgg 428
          |||||||||||||||||||||||||||||| ||||| |||||||||||||| ||  |||
Db    254 TGCGGCAGCTCAACATGTATGGCTTCCGGAAGGTGGTGCACATCGAGCAGGGAGGGTTGG 313

Qy    429 tcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtggccaggagc 488
          | |||||||||  ||||||| |||||||||||||| |  ||| | ||||||||||||||
Db    314 TGAAGCCAGAGAAGGACGACACTGAGTTCCAGCACCCCTATTTCATCCGTGGCCAGGAGC 373

Qy    489 agctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagaca 548
          | ||||| |||||||||||||||||||| ||||| ||||||| | | ||||  || ||||
Db    374 ACCTCCTGGAGAACATCAAGAGGAAAGTCACCAGCGTGTCCAGCATAAAGAACGAGGACA 433

Qy    549 taaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaagggga 608
          | ||| |  | || |||| |||||||||||||||||| ||| | ||| ||||||| || |
Db    434 TCAAGGTGAGGCAAGACAACGTCACCAAGCTGCTGACCGACATCCAGGTGATGAAAGGAA 493

Qy    609 agcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggctctgtggc 668
          |||||||| |||||||||||||||  | |||||||||||||||| ||||| ||||||
Db    494 AGCAGGAGAGCATGGACTCCAAGCTGATAGCCATGAAGCATGAGAACGAGGCGCTGTGGA 553

Qy    669 gggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctca 728
          ||||||||||||||  |||||||||| |||||||| ||||| || ||||| |||||||
Db    554 GGGAGGTGGCCAGCCTAAGGCAGAAGCACGCCCAGCAGCAGAAGGTGGTCAATAAGCTCA 613

Qy    729 ttcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagagaaagatcc 788
          | |||||| |||| ||| ||||||| || ||||| ||||| |||||||||||| |||||||
Db    614 TCCAGTTCTTGATTTCATTGGTGCAATCCAACCGCATCCTCGGGGTGAAGAGGAAGATCC 673

Qy    789 ccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccggcagttct 848
          ||||||| |||||||||  | ||||| |||||||||| || || ||||| |||| ||
Db    674 CCCTGATGTTGAACGACAGCAGTTCAGCTCATTCCATGCCGAAATACAGCCGCCAGTACT 733

Qy    849 ccctggagcacgtccacggctcgggcccctactcggccccctccccagcctacagcagct 908
          ||||||||||||||||||||||   ||| ||| | ||  | ||||||| |||||| |||
```

```
Db    734 CCCTGGAGCACGTCCACGGCTCATCCCCATACGCAGCTTCTTCCCCAGCGTACAGCGGCT 793

Qy    909 ccagcctctacgcccctgatgctgtggccagctctggacccatcatctccgacatcaccg 968
          ||| | ||||| |||| ||| |     ||| ||| || |||||||||| ||| | || |
Db    794 CCAACATCTACTCCCCCGATTCATCCACCAACTCCGGCCCCATCATCTCGGACGTGACGG 853

Qy    969 agctggctcctgccagccccatggcctcccccggcgggagcatagacgagaggcccctat 1028
          ||||||| |   |||||||    || || ||| | || ||| | || ||||||
Db    854 AGCTGGCCCAGTCCAGCCCTTCAGCATCTCCCAGTGGCAGCCTGGATGAGAGG------- 907

Qy   1029 ccagcagccccctggtgcgtgtcaaggaggagcccccagcccgcctcagagcccccggg 1088
            ||||||||  |||| ||  |||| ||||||||||||||||||  | |  ||||||  ||
Db    907 --AGCAGCCCTGTGGTCCGCATCAAAGAGGAGCCCCCCAGCCCCTCCCGCAGCCCCAAGG 964

Qy   1089 tagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtccccgaccgccc 1148
            | ||| | ||  ||  |   |  | |  |||  |||
Db    965 AGAACGAGCCCAGCACCACGACGGCCGCGGCCGGCAACA---------------------- 1004

Qy   1149 tcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagccctcacgg 1208
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1004 ---------------------------------------------------------- 1004

Qy   1209 acgccaggggccacacggacaccgagggccggcctccctccccccccgcccacctccaccc 1268
                                                  | ||  ||   |    |||
Db   1004 ------------------------------------GCACCGAGCAGCCTCAGCCCC 1024

Qy   1269 ctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgaccacttggatg 1328
            |||||||||||||||| |||||||| |||||||| || || |  |||||| ||||
Db   1025 AGGAAAAGTGCCTCAGCGTTGCCTGCCTCGACAAGAACGAACTGAACGACCACCTGGACA 1084

Qy   1329 ctatggactccaacctggataacctgcagaccatgctgagcagccacggcttcagcgtgg 1388
          | || ||||||||||||||| ||||||||||||||||||||  ||||||||||||| |
Db   1085 CCATCGACTCCAACCTGGACAACCTGCAGACCATGCTGAGCACGCACGGCTTCAGCGTCG 1144

Qy   1389 acaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacatgagcctgc 1448
          ||||||  |||||||||||||||| ||||||||||  |||| ||  | || |||| ||| |
Db   1145 ACACCACCGCCCTGCTGGACCTCTTCAGCCCCTCCATGACGGTTACGGATATGAACCTCC 1204

Qy   1449 ctgaccttgacagcagcctggccagtatccaagagctcctgtctccccaggagcccccca 1508
          | ||||| |||||||||| ||||| ||||| || || || ||  | |||||||  |  |
Db   1205 CCGACCTGGACAGCAGCCTCGCCAGCATCCAGGATCTGCTCTCGTCGCAGGAGCAGCAGA 1264

Qy   1509 ggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgcactacacag 1568
           |||  | ||||| ||           | ||  |||||||||||| |||| |||||||| |
Db   1265 AGCCCTCGGAGGCCGATGCGGCAGCAGCAGACACAGGGAAGCAGTTGGTTCACTACACGG 1324

Qy   1569 cgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacgacctgccgg 1628
          | ||||| || |||||| ||||  |  || | ||||||   || |||   ||||| ||
Db   1325 CTCAGCCCCTCTTCCTGGTGGATTCGAGCGCGGTGGACGTGGGCAGCGGAGACCTCCCCA 1384

Qy   1629 tgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcgccgaggacc 1688
          |  | || ||| |||| |||||||||||| ||| |||| |||||||  | |  ||| ||||
Db   1385 TCTTCTTCGAGTTGGGGGAGGGCTCCTATTTCACCGACGGGGACGAATACAACGAAGACC 1444

Qy   1689 ccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccccactgtctcct 1748
          |||||||||||||| ||  | ||  | ||||  |||||| |||||||||||||| || ||||
Db   1445 CCACCATCTCCCTCCTATCCGGGACTGAGCAACCCAAACCCAAGGACCCCACCGTGTCCT 1504

Qy   1749 agaggc 1754
          | || |
Db   1505 AAAGCC 1510
```

```
RESULT   6
AF205589/c
LOCUS       AF205589   131973 bp    DNA             HTG       07-DEC-1999
DEFINITION  Homo sapiens chromosome 8 clone BAC 393G12, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AF205589
VERSION     AF205589.1  GI:6531668
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 131973)
  AUTHORS   Polley,A., Wen,G., Baumgart,C., Dette,M., Jahn,N., Schilhabel,M.,
            Menzel,U. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     1-16995: contig of 16995 bp;  16995-16996: gap ofunknown size;
            16996-28004: contig of 11009 bp;  28004-28005: gap of unknownsize;
            28005-43979: contig of 15975 bp;  43979-43980: gap of unknown size;
             43980-67713: contig of 23734 bp;  67713-67714: gap of unknown
            size;  67714-77798: contig of 10085 bp;  77798-77799: gap of
            unknown size;  77799-92762: contig of 14964 bp;  92762-92763: gap
            of unknown size;  92763-98839: contig of 6077 bp;  98839-98840: gap
            of unknown size;  98840-104153: contig of 5314 bp;  104153-104154:
            gap of unknown size;  104154-108012: contig of 3859 bp;
            108012-108013: gap of unknown size;  108013-111572: contig of 3560
            bp;  111572-111573: gap of unknown size;  111573-113894: contig of
            2322 bp;  113894-113895: gap of unknown size;  113895-131973:
            contig of 18079 bp;.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .131973
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="BAC 393G12"
                     /note="seed clone"
     exon            complement(213. .614)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            240. .541
                     /note="MZEF"
                     /evidence=not_experimental
     exon            240. .541
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            240. .541
                     /note="GENSCAN"
     exon            240. .547
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            760. .931
                     /note="MZEF"
                     /evidence=not_experimental
     exon            760. .931
                     /note="GENSCAN"
     exon            760. .871
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            763. .891
                     /note="XPOUND"
                     /evidence=not_experimental
     repeat_region   1011. .1136
                     /rpt_family="FLAM_A"
     misc_feature    1138. .1441
                     /note="CpGisland: CpG > 5%"
     repeat_region   1144. .1439
                     /rpt_family="AluY"
     exon            1574. .1759
                     /note="GENSCAN"
     exon            2040. .2234
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            2059. .2234
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            2059. .2234
                     /note="GENSCAN"
     exon            2182. .2234
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   2394. .2701
                     /rpt_family="AluSg"
     repeat_region   2707. .3012
                     /rpt_family="AluY"
     exon            3368. .3441
                     /note="GENSCAN"
     exon            3982. .4115
                     /note="GENSCAN"
     exon            4019. .4115
                     /note="XPOUND"
                     /evidence=not_experimental
     misc_feature    5092. .6790
                     /note="CpGisland: CpG > 5%"
     exon            5107. .5159
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            5162. .5376
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            5183. .5376
                     /note="GENSCAN"
     exon            5183. .5376
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   5431. .5735
                     /rpt_family="AluY"
     misc_feature    5774. .6055
                     /note="CpGisland: CpG > 10%"
     exon            5955. .6030
                     /note="GENSCAN"
     exon            complement(6254. .6367)
                     /note="XPOUND"
                     /evidence=not_experimental
     repeat_region   6573. .6730
                     /rpt_family="L2"
     repeat_region   complement(7193. .7499)
                     /rpt_family="AluSx"
     exon            7920. .7955
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            7920. .7955
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            8311. .8339
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8311. .8339
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8311. .8339
                     /note="GENSCAN"
     exon            8369. .8493
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            8404. .8493
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8456. .8493
                     /note="GENSCAN"
     exon            8456. .8493
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8731. .8836
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8738. .8836
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            8747. .8836
                     /note="MZEF"
```

```
                     /evidence=not_experimental
     exon            8747. .8836
                     /note="GENSCAN"
     exon            9468. .9573
                     /note="MZEF"
                     /evidence=not_experimental
     exon            9468. .9573
                     /note="GENSCAN"
     exon            9479. .9573
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            9479. .9573
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            9700. .9801
                     /note="MZEF"
                     /evidence=not_experimental
     exon            9700. .9801
                     /note="GENSCAN"
     exon            9700. .9801
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            9700. .9801
                     /note="GRAIL"
                     /evidence=not_experimental
     misc_feature    10188. .10698
                     /note="CpGisland: CpG > 5%"
     exon            10203. .10304
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            complement(10241. .10371)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10251. .10304
                     /note="GENSCAN"
     exon            10251. .10334
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10251. .10304
                     /note="MZEF"
                     /evidence=not_experimental
     exon            10345. .10478
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10387. .10475
                     /note="XPOUND"

  Query Match            27.8%;  Score 599;  DB 44;  Length 131973;
  Best Local Similarity  95.0%;  Pred. No. 8.4e-94;
  Matches  630;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    1;

Qy   1475 atccaagagctcctgtctccccaggagccccccaggcctcccgaggcagagaacagcagc 1534
          | |||||  |  | | | || |    ||||||| |  || | |      || | |  || |
Db  50579 AGCCAAGTCCCACCGGCCCCACCTTTGCCCCCAACCCCCCACCGCCTTGACACCCCCACC 50520

Qy   1535 ccggattcagggaagcagctggtgcactacacagcgcagccgctgttcctgctggacccc 1594
          || |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50519 CCCG---CAGGGAAGCAGCTGGTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCC 50463

Qy   1595 ggctccgtggacaccgggagcaacgacctgccggtgctgtttgagctgggagagggctcc 1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50462 GGCTCCGTGGACACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCC 50403

Qy   1655 tacttctccgaaggggacggcttcgccgaggaccccaccatctccctgctgacaggctcg 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50402 TACTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCG 50343

Qy   1715 gagcctcccaaagccaaggaccccactgtctcctagaggccccggaggagctgggccagc 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50342 GAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGC 50283

Qy   1775 cgcccacccccacccccagtgcagggctggtcttggggaggcagggcagcctcgcggtct 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  50282 CGCCCACCCCCACCCCCAGTGCAGGGCTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCT 50223

Qy   1835 tgggcactggtgggtcggccgccatagccccagtaggacaaacgggctcgggtctgggca 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50222 TGGGCACTGGTGGGTCGGCCGCCATAGCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCA 50163

Qy   1895 gcacctctggtcaggagggtcaccctggcctgccagtctgccttcccccaaccccgtgtc 1954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50162 GCACCTCTGGTCAGGAGGGTCACCCTGGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTC 50103

Qy   1955 ctgtggtttggttggggcttcacagccacacctggactgaccctgcaggttgttcatagt 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50102 CTGTGGTTTGGTTGGGGCTTCACAGCCACACCTGGACTGACCCTGCAGGTTGTTCATAGT 50043

Qy   2015 cagaattgtattttggatttttacacaactgtcccgttccccgctccacagagatacaca 2074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50042 CAGAATTGTATTTTGGATTTTTACACAACTGTCCCGTTCCCCGCTCCACAGAGATACACA 49983

Qy   2075 gatatatacacacagtggatggacggacaagacaggcagagatctataaacagacaggct 2134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  49982 GATATATACACACAGTGGATGGACGGACAAGACAGGCAGAGATCTATAAACAGACAGGCT 49923

Qy   2135 cta 2137
          |||
Db  49922 CTA 49920


RESULT   7
XELHSF
LOCUS       XELHSF       1657 bp    mRNA            VRT       16-MAY-1996
DEFINITION  Xenopus laevis heat shock factor 1 (XHSF1) mRNA, complete cds.
ACCESSION   L36924
VERSION     L36924.1  GI:558067
KEYWORDS    heat shock factor.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
            Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
            Xenopus.
REFERENCE   1  (bases 1 to 1657)
  AUTHORS   Stump,D.G., Landsberger,N. and Wolffe,A.P.
  TITLE     The cDNA encoding Xenopus laevis heat-shock factor 1 (XHSF1):
            nucleotide and deduced amino-acid sequences, and properties of the
            encoded protein
  JOURNAL   Gene 160 (2), 207-211 (1995)
  MEDLINE   95369690
REFERENCE   2  (bases 1 to 1657)
  AUTHORS   Landsberger,N. and Wolffe,A.P.
  TITLE     Role of chromatin and Xenopus laevis heat shock transcription
            factor in regulation of transcription from the X. laevis hsp70
            promoter in vivo
  JOURNAL   Mol. Cell. Biol. 15 (11), 6013-6024 (1995)
  MEDLINE   96025982
FEATURES             Location/Qualifiers
     source          1. .1657
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /cell_line="A6"
     5'UTR           1. .226
     mRNA            1. .1657
     gene            227. .1582
                     /gene="XHSF1"
     CDS             227. .1582
                     /gene="XHSF1"
                     /codon_start=1
                     /product="heat shock factor"
                     /protein_id="AAA99999.1"
                     /db_xref="GI:558068"
                     /translation="MDPHGTCGGSNVPAFLAKLWTLVEDPDTDPLICWSPEGNSFHVF
                     DQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVHIEQGGLVKPERDDTEFQHPYFI
                     RGQEQLLENIKRKVNTMSATKSDEVKVRQDSVGKLISDVQSMKGKQESIDGRLLSMKH
                     ENEALWREVASLRQKHTQQQKVVNKLIQFLVSLVQSNRILGVKRKIPLMLNDSSTGHS
                     PPKYSRQYSLEHVPSSTSYPVSGFTDSSAGPIISDVTELPESSPSPSPCPSLEASPSP
```

```
                 VILIKTEPLTPSQSPEQSPAPPKLDDTPISPSTFIDSILLETETSVCPGGNKNDEMSE
                 SHPPEPCLSVACLDNISLSRQMSEVSRLFPTSCSSVPGRAEPPGLDMAVAELNDHVDN
                 IDFNLDTLQNLLNGQSFSVDTSALMDLFSPSLGIPDLSLPDPDSSLASVSSTPIYTCV
                 V"
    3'UTR        1583. .1657
BASE COUNT      400 a    467 c    409 g    381 t
ORIGIN
```

```
  Query Match            22.1%;  Score 476;  DB 4;  Length 1657;
  Best Local Similarity  61.5%;  Pred. No. 1.8e-72;
  Matches  810;  Conservative    0;  Mismatches  485;  Indels   23;  Gaps    2;

Qy    178 ccccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgt 237
          || |||  |    |||  |||||| || || || |||||  |||||||  |||||  | ||
Db    235 CCACGGGACTTGTGGGGGCAGCAATGTACCTGCTTTCCTTGCCAAGCTCTGGACTTTGGT 294

Qy    238 gagcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgt 297
          |  ||||| ||||| ||| |  | ||||||||||||||    |||||||| |||| ||
Db    295 GGAAGACCCAGACACAGACCCTTTAATCTGCTGGAGCCCAGAAGGGAACAGTTTCCATGT 354

Qy    298 gttcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacat 357
          ||| |||||||| |||||||||| ||||| || || ||||| ||||||||||||||||
Db    355 GTTTGACCAGGGACAGTTTGCCAAAGAGGTTCTTCCAAAGTATTTCAAGCACAACAACAT 414

Qy    358 ggccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagca 417
          |||||||||| || ||||| ||||| |||||||| || ||||| || ||||| || || ||
Db    415 GGCCAGCTTTGTACGGCAACTCAATATGTATGGGTTTCGGAAGGTTGTCCATATTGAACA 474

Qy    418 gggcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcg 477
          || |||||||| || || || ||||| ||||| |||||||||||||| | |||| | ||
Db    475 AGGGGGCCTGGTGAAACCTGAAAGAGATGACACAGAGTTCCAGCACCCCTACTTCATCCG 534

Qy    478 tggccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaa 537
          ||| |||||||| || || |||||||||||| | ||||| | ||   |||| |    ||
Db    535 TGGTCAGGAGCAACTTCTGGAGAACATCAAGCGCAAAGTCAACACCATGTCTGCAACAAA 594

Qy    538 gagtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagct 597
          ||| || ||  | ||| | || |||||||| ||    |||| || | |  | || || |||
Db    595 GAGCGACGAGGTGAAGGTGCGACAGGACAGTGTTGGCAAGTTGATATCAGATGTACAGAG 654

Qy    598 gatgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatga 657
          ||||||| |||||||| |||| ||   |  ||| ||  ||||||| |||||||| ||
Db    655 CATGAAGGGCAAGCAGGAGAGCATTGATGGCCGGCTACTCTCCATGAAACATGAGAACGA 714

Qy    658 ggctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgt 717
          ||| |||||| |||||||| || |||||  |||||||||  | ||||| |||||||| ||
Db    715 GGCCCTGTGGAGGGAGGTTGCGAGCCTGAGGCAGAAGCACACTCAGCAGCAGAAAGTGGT 774

Qy    718 caacaagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaa 777
          ||| |||||||| |||||||| ||||| |||| |||||||||||||||||||||||| ||
Db    775 CAATAAGCTCATCCAGTTCCTGGTCTCATTGGTTCAGTCAAACCGGATCCTGGGGGTAAA 834

Qy    778 gagaaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatag 837
          ||| || |||||||| |||||||| |||||| |  |||  || |||   |||||||| ||
Db    835 GAGGAAAATCCCCCTCATGCTGAATGACAGTAGTACAGGCCACTCCCCTCCCAAGTACAG 894

Qy    838 ccggcagttctccctggagcacgtccacggctcgggcccctactcggccccctccccagc 897
          || |||| |||  |||||||||| | | ||     | ||||| | | | |       |
Db    895 TCGTCAGTACTCATTGGAGCACGTGCCCAGCAGCACCTCCTACCCAGTCTCGGGATTCAC 954

Qy    898 ctacagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctc 957
            ||  |    |  | | |       ||||||    ||    | | |   ||| |  |||
Db    955 AGACTCCTCGGCAGGACCCATTATATCTGATGTGACGGAACTCCCAGAGTCCAGCCCCTC 1014

Qy    958 cgacatcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacga 1017
          |     |  |    || | |     ||   |  | ||    || | |         | ||
Db   1015 C----CCCTCTCCGTGTCCCAGTTTAGAGGCCAGCCCCAGTCCAGTGATTCTGATCAAGA 1070

Qy   1018 gaggcccctatccagcagccccctggtgcgtgtcaaggaggagccccccagcccgcctca 1077
            |  |||    |  || |||   |    | |   ||       |||||    |    |
Db   1071 CGGAGCCCCTCACTCCATCCCAGAGCCCTGAGCAGAGTCCTGCGCCCCCTAAACTGGACG 1130
```

```
Qy   1078 gagcccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtc 1137
            ||||    | |          ||   | | |   ||| | |  |||     |  ||
Db   1131 ACACCCCAATATCGCCCTCCACTTTCATAGACTCAATTCTGCTGGAGACTGAGACCAGTG 1190

Qy   1138 cccgacc------------------gccctcattgactccatcctgcgggagagtgaac 1178
            | ||                  |   |    || ||   ||  |  |||      |
Db   1191 TGTGTCCCGGGGGCAATAAAAATGATGAAATGTCAGAGTCGCACCCCCCTGAGCCCTGCC 1250

Qy   1179 ctgcccccgcctccgtcacagccctcacggacgccaggggccacacggacaccgagggcc 1238
                |||||  |      | || |      | |         | ||   |    | |
Db   1251 TGAGTGTCGCCTGTCTAGACAACATCTCTTTGAGCCGTCAGATGTCTGAGGTCTCTCGTC 1310

Qy   1239 ggcctccctccccccccgcccacctccacccctgaaaagtgcctcagcgtagcctgcctgg 1298
            |   ||| |  ||    || |     |       |  |   | |  ||         |||
Db   1311 TGTTCCCCACGTCCTGTTCCTCAGTGCCGGGCAGAGCGGAGCCCCCCGGGTTAGATATGG 1370

Qy   1299 acaagaatgagctcagtgaccacttggatgctatggactccaacctggataacctgcaga 1358
             |   ||||| | ||||||  ||||    || |||| |||||||||| | ||||||||
Db   1371 CTGTGGCGGAGCTAAATGACCATGTGGACAACATTGACTTCAACCTGGACACCCTGCAGA 1430

Qy   1359 ccatgctgagcagccacggcttcagcgtggacaccagtgccctgctggacctgttcagcc 1418
            | | || |   |||| |||||||| || || |||||||| ||  |||| || ||||| |
Db   1431 ACCTTCTCAATGGCCAGAGCTTCAGTGTAGATACCAGTGCTCTCATGGATCTCTTCAGTC 1490

Qy   1419 cctcggtgaccgtgcccgacatgagcctgcctgaccttgacagcagcctggccagtat 1476
           | ||  ||    | || ||  ||  ||| || ||||  |||  |||||| |||||| |
Db   1491 CGTCCTTGGGAATCCCAGATCTGTCCCTCCCAGACCCCGACTCCAGCCTCGCCAGTGT 1548
```

```
RESULT   8
D87673
LOCUS       D87673       1555 bp    mRNA            PRI       07-FEB-1999
DEFINITION  Homo sapiens mRNA for heat shock transcription factor 4, complete
            cds.
ACCESSION   D87673
VERSION     D87673.1  GI:1813425
KEYWORDS    hHSF4; heat shock transcription factor 4.
SOURCE      Homo sapiens adult heart cDNA to mRNA, clone:phHSF4-7a.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1555)
  AUTHORS   Nakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Akira
            Nakai, Chest Disease Research Institute, Department of Cell
            Biology; Sakyo-ku, Kawahara-chyo 53, Kyoto, Kyoto 606-01, Japan
            (E-mail:nakai@chest.kyoto-u.ac.jp, Tel:075-751-3846,
            Fax:075-752-9017)
REFERENCE   2  (bases 1 to 1555)
  AUTHORS   Nakai,A., Tanabe,M., Kawazoe,Y., Inazawa,J., Morimoto,R. and
            Nagata,K.
  TITLE     HSF4, a new member of the human heat shock factor gene family which
            lacks properties of a transcriptional activator
  JOURNAL   Unpublished (1997)
REFERENCE   3  (sites)
  AUTHORS   Nakai,A., Tanabe,M., Kawazoe,Y., Inazawa,J., Morimoto,R.I. and
            Nagata,K.
  TITLE     HSF4, a new member of the human heat shock factor family which
            lacks properties of a transcriptional activator
  JOURNAL   Mol. Cell. Biol. 17 (1), 469-481 (1997)
  MEDLINE   97127404
FEATURES             Location/Qualifiers
     source          1. .1555
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="phHSF4-7a"
                     /dev_stage="adult"
                     /tissue_type="heart"
     gene            5. .1396
                     /gene="hHSF4"
     CDS             5. .1396
                     /gene="hHSF4"
```

```
                     /function="it may regulate transcription of heat shock
                     genes"
                     /codon_start=1
                     /product="heat shock transcription factor 4"
                     /protein_id="BAA13433.1"
                     /db_xref="GI:1813426"
                     /translation="MVQEAPAALPTEPGPSPVPAFLGKLWALVGDPGTDHLIRWSPSG
                     TSFLVSDQSRFAKEVLPQYFKHSNMASFVRQLNMYGFRKVVSIEQGGLLRPERDHVEF
                     QHPSFVRGREQLLERVRRKVPALRGDDGRWRPEDLGRLLGEVQALRGVQESTEARLRE
                     LRQQNEILWREVVTLRQSHGQQHRVIGKLIQCLFGPLQAGPSNAGGKRKLSLMLDEGS
                     SCPTPAKFNTCPLPGALLQDPYFIQSPSTYSLSQRQIWALALTGPGAPSSLTSQKTLH
                     PLRGPGFLPPVMAGAPPPLPVAVVQAILEGKGSFSPEGPRNAQQPEPGDPREIPDRGP
                     LGLESGDRSPESLLPPMLLQPPQESVEPAGPLDVLGPSLQGREWTLMDLDMELSLMQP
                     LVPERGEPELAVKGLNSPSPGKDPTLGAPLLLDVQAALGGPALGLPGALTIYSTPESR
                     TASYLGPEASPSP"
BASE COUNT      290 a    503 c    487 g    275 t
ORIGIN
```

```
  Query Match             12.9%;  Score 277.4;  DB 9;  Length 1555;
  Best Local Similarity   64.7%;  Pred. No. 1.7e-38;
  Matches  465;  Conservative    0;  Mismatches  241;  Indels   13;  Gaps    3;

Qy     153 tgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggcct 212
           || |  ||   |  |   | | |   |||   | | | || ||||||  ||| || ||||
Db       6 TGGTGCAGGAAGCGCCAGCTGCGCTGCCCACGGAGCCAGGCCCCAGCCCCGTGCCTGCCT 65

Qy     213 tcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctgga 272
           ||||  ||||||| ||| | || ||| | ||||| | ||| |||   || ||| |||||| 
Db      66 TCCTCGGCAAGCTATGGGCGCTGGTGGGGGACCCAGGCACAGACCACCTGATCCGCTGGA 125

Qy     273 gcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgctgc 332
           |||||||||||| ||| |||| |||   ||||||| |||  || ||||||| |||||||
Db     126 GCCCGAGCGGGACCAGTTTCCTCGTAAGCGACCAGAGCCGTTTCGCCAAGGAAGTGCTGC 185

Qy     333 ccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtatggct 392
           || |||| |||||||| | |||||||| |||||||||| || ||||||||||| || |
Db     186 CCCAGTATTTCAAGCATAGCAACATGGCGAGCTTCGTGCGCCAACTCAACATGTACGGTT 245

Qy     393 tccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgacacgg 452
           | ||||| |||||   ||||||||||||||||||| | | ||| ||| | ||| ||   |
Db     246 TTCGGAAGGTGGTGAGCATCGAGCAGGGCGGCCTGCTTAGGCCGGAGCGCGACCACGTCG 305

Qy     453 agttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaagagga 512
           ||||||||||||  ||||| |||| ||||  |||||||| || |||  | |   | | |
Db     306 AGTTCCAGCACCCGAGCTTCGTGCGCGGCCGCGAGCAGCTACTGGAGCGCGTGCGGCGCA 365

Qy     513 aagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgtca 572
           |          ||| || | |||    | || |||         |||| |||   | |
Db     366 AG---------GTGCCCGCGCTGCGCGGCGACGACGGCCGCTGGCGCCCGGAGGACCTGG 416

Qy     573 ccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaagc 632
               || |||   || ||||||    ||  |||   ||||||| |||  ||  |   ||
Db     417 GTCGACTACTGGGCGAGGTGCAGGCTTTGCGGGGAGTGCAGGAGAGCACCGAGGCGCGGC 476

Qy     633 tcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagccttcggcaga 692
           | | ||   | | |||  |||| |||    ||||||||||||||  |  ||||||||||
Db     477 TGCGGGAGCTCAGGCAGCAGAACGAGATCTTGTGGCGGGAGGTGGTGACACTTCGGCAGA 536

Qy     693 agcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggtgc 752
             || |  ||||| ||    ||| |   |||||| || ||||  ||  |    |   | |
Db     537 GCCACGGTCAGCAGCACCGGGTCATTGGCAAGCTGATCCAGTGTCTCTTTGGGCCACTTC 596

Qy     753 agtcaaaccggatc---ctgggggtgaagagaaagatccccctgatgctgaacgacagtg 809
           || |    | || |      || |  |||||||||  |  ||||||||||| | ||  |
Db     597 AGGCGGGGCCGAGCAATGCAGGAGGCAAGAGAAAGCTGTCCCTGATGCTGGATGAGGGGA 656

Qy     810 gctca-gcacattccatgcccaagtatagccggcagttctccctggagcacgtccacgg 867
           ||||| || ||   | |||| |  |  |  |       ||| ||||| || | ||  | |
Db     657 GCTCATGCCCAACACCTGCCAAGTTCAACACCTGCCCTCTACCTGGTGCCCTTCTGCAG 715

RESULT   9
```

```
AB029348
LOCUS       AB029348     1577 bp    mRNA            PRI       23-SEP-1999
DEFINITION  Homo sapiens mRNA for transcription factor HSF4b isoform, complete
            cds.
ACCESSION   AB029348
VERSION     AB029348.1  GI:5921134
KEYWORDS    transcription factor HSF4b isoform.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Tanabe,M., Sasai,N., Nagata,K., Liu,X.D., Liu,P.C., Thiele,D.J. and
            Nakai,A.
  TITLE     The mammalian HSF4 gene generates both an activator and a repressor
            of heat shock genes by alternative splicing
  JOURNAL   J. Biol. Chem. 274 (39), 27845-27856 (1999)
  MEDLINE   99419073
REFERENCE   2  (bases 1 to 1577)
  AUTHORS   Nakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) to the DDBJ/EMBL/GenBank databases. Akira
            Nakai, Institute for Frontier Medical Sciences, Department of
            Molecular and Cell Biology; Sakyo-ku, Kyoto 606-8397, Japan
            (E-mail:nakai@frontier.kyoto-u.ac.jp, Tel:81-75-751-4638,
            Fax:81-75-752-9017)
FEATURES             Location/Qualifiers
     source          1. .1577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             5. .1486
                     /codon_start=1
                     /product="transcription factor HSF4b isoform"
                     /protein_id="BAA84582.1"
                     /db_xref="GI:5921135"
                     /translation="MVQEAPAALPTEPGPSPVPAFLGKLWALVGDPGTDHLIRWSPSG
                     TSFLVSDQSRFAKEVLPQYFKHSNMASFVRQLNMYGFRKVVSIEQGGLLRPERDHVEF
                     QHPSFVRGREQLLERVRRKVPALRGDDGRWRPEDLGRLLGEVQALRGVQESTEARLRE
                     LRQQNEILWREVVTLRQSHGQQHRVIGKLIQCLFGPLQAGPSNAGGKRKLSLMLDEGS
                     SCPTPAKFNTCPLPGALLQDPYFIQSPLPETNLGLSPHRARGPIISDIPEDSPSPEGT
                     RLSPSSDGRREKGLALLKEEPASPGGDGEAGLALAPNECDFCVTAPPPLPVAVVQAIL
                     EGKGSFSPEGPRNAQQPEPGDPREIPDRGPLGLESGDRSPESLLPPMLLQPPQESVEP
                     AGPLDVLGPSLQGREWTLMDLDMELSLMQPLVPERGEPELAVKGLNSPSPGKDPTLGA
                     PLLLDVQAALGGPALGLPGALTIYSTPESRTASYLGPEASPSP"
BASE COUNT      300 a    494 c    523 g    260 t
ORIGIN
```

```
  Query Match             12.9%;  Score 277.4;  DB 10;  Length 1577;
  Best Local Similarity   64.7%;  Pred. No. 1.7e-38;
  Matches  465;  Conservative    0;  Mismatches  241;  Indels   13;  Gaps    3;

Qy     153 tgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggcct 212
           || |  ||   |  |   | | |   |||   | | | || ||||||  ||| || ||||
Db       6 TGGTGCAGGAAGCGCCAGCTGCGCTGCCCACGGAGCCAGGCCCCAGCCCCGTGCCTGCCT 65

Qy     213 tcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctgga 272
           ||||  ||||||| ||| | || ||| | ||||| | ||| |||   || ||| |||||| 
Db      66 TCCTCGGCAAGCTATGGGCGCTGGTGGGGGACCCAGGCACAGACCACCTGATCCGCTGGA 125

Qy     273 gcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgctgc 332
           |||||||||||| ||| |||| |||   ||||||| |||  || ||||||| |||||||
Db     126 GCCCGAGCGGGACCAGTTTCCTCGTAAGCGACCAGAGCCGTTTCGCCAAGGAAGTGCTGC 185

Qy     333 ccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtatggct 392
           || |||| |||||||| | |||||||| |||||||||| || ||||||||||| || |
Db     186 CCCAGTATTTCAAGCATAGCAACATGGCGAGCTTCGTGCGCCAACTCAACATGTACGGTT 245

Qy     393 tccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgacacgg 452
           | ||||| |||||   ||||||||||||||||||| | | ||| ||| | ||| ||   |
Db     246 TTCGGAAGGTGGTGAGCATCGAGCAGGGCGGCCTGCTTAGGCCGGAGCGCGACCACGTCG 305

Qy     453 agttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaagagga 512
```

```
Db    306 AGTTCCAGCACCCGAGCTTCGTGCGCGGCCGCGAGCAGCTACTGGAGCGCGTGCGGCGCA 365

Qy    513 aagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgtca 572
Db    366 AG---------GTGCCCGCGCTGCGCGGCGACGACGGCCGCTGGCGCCCGGAGGACCTGG 416

Qy    573 ccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaagc 632
Db    417 GTCGACTACTGGGCGAGGTGCAGGCTTTGCGGGGAGTGCAGGAGAGCACCGAGGCGCGGC 476

Qy    633 tcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagccttcggcaga 692
Db    477 TGCGGGAGCTCAGGCAGCAGAACGAGATCTTGTGGCGGGAGGTGGTGACACTTCGGCAGA 536

Qy    693 agcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggtgc 752
Db    537 GCCACGGTCAGCAGCACCGGGTCATTGGCAAGCTGATCCAGTGTCTCTTTGGGCCACTTC 596

Qy    753 agtcaaaccggatc---ctgggggtgaagagaaagatccccctgatgctgaacgacagtg 809
Db    597 AGGCGGGGCCGAGCAATGCAGGAGGCAAGAGAAAGCTGTCCCTGATGCTGGATGAGGGGA 656

Qy    810 gctca-gcacattccatgcccaagtatagccggcagttctccctggagcacgtccacgg 867
Db    657 GCTCATGCCCAACACCTGCCAAGTTCAACACCTGCCCTCTACCTGGTGCCCTTCTGCAG 715
```

RESULT 10
AB029350
LOCUS AB029350 1591 bp mRNA ROD 23-SEP-1999
DEFINITION Mus musculus mRNA for transcription factor HSF4a isoform, complete cds.
ACCESSION AB029350
VERSION AB029350.1 GI:5921138
KEYWORDS transcription factor HSF4a isoform.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Tanabe,M., Sasai,N., Nagata,K., Liu,X.D., Liu,P.C., Thiele,D.J. and Nakai,A.
TITLE The mammalian HSF4 gene generates both an activator and a repressor of heat shock genes by alternative splicing
JOURNAL J. Biol. Chem. 274 (39), 27845-27856 (1999)
MEDLINE 99419073
REFERENCE 2 (bases 1 to 1591)
AUTHORS Nakai,A.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) to the DDBJ/EMBL/GenBank databases. Akira Nakai, Institute for Frontier Medical Sciences, Department of Molecular and Cell Biology; Sakyo-ku, Kyoto 606-8397, Japan (E-mail:nakai@frontier.kyoto-u.ac.jp, Tel:81-75-751-4638, Fax:81-75-752-9017)
FEATURES Location/Qualifiers
source 1. .1591
/organism="Mus musculus"
/db_xref="taxon:10090"
CDS 1. .1389
/codon_start=1
/product="transcription factor HSF4a isoform"
/protein_id="BAA84584.1"
/db_xref="GI:5921139"
/translation="MQEAPAALPTEPGPSPVPAFLGKLWALVGDPGTDHLIRWSPSGT
SFLVSDQSRFAKEVLPQYFKHSNMASFVRQLNMYGFRKVVSIEQGGLLRPERDHVEFQ
HPSFVRGREQLLERVRRKVPALRGDDSRWRPEDLSRLLGEVQALRGVQESTEARLQEL
RQQNEILWREVVTLRQSHSQQHRVIGKLIQCLFGPLQTGPSSTGAKRKLSLMLDEGSA
CSASAKFNACPVSGALLQDPYFIQSPSPCSPSQRPRWASALTGPEGPSSLTSQKILHL
LKDTGFLPPVVAGAPPPLPVAVVQAILEGKGSYSPEGPRSVQQPEPRGPREVPDRGTL
GLDRGNRSPESLLPPMLLRPAPETLEPVAPVDVLGPSLHGREWTLMDLDMELSLMQPL
APETDEAELTVKELNSSGVGKDHTLGTPLMLDVQADLEGAALSVPGALTLYNVTESNA
SYLDPGASPSSP"

```
BASE COUNT 347 a 495 c 458 g 291 t
ORIGIN

Query Match 12.0%; Score 259.6; DB 12; Length 1591;
Best Local Similarity 65.4%; Pred. No. 1.9e-35;
Matches 417; Conservative 0; Mismatches 209; Indels 12; Gaps 2;

Qy    179 cccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtg 238
Db     25 CCCACGGAGCCAGGCCCCAGCCCGGTACCTGCCTTCCTCGGCAAGCTATGGGCGCTGGTA 84

Qy    239 agcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtg 298
Db     85 GGCGACCCAGGCACCGACCACCTCATCCGCTGGAGCCCGAGTGGCACCAGCTTCCTCGTA 144

Qy    299 ttcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatg 358
Db    145 AGTGATCAGAGCCGCTTCGCCAAGGAAGTGCTGCCCCAGTACTTCAAGCACAGCAACATG 204

Qy    359 gccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcag 418
Db    205 GCGAGCTTTGTTCGCCAACTCAACATGTATGGTTTTCGGAAGGTGGTGAGCATCGAGCAA 264

Qy    419 ggcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgt 478
Db    265 GGTGGCCTGCTCAGACCAGAGCGTGACCACGTTGAGTTTCAGCATCCGAGCTTCGTGCGA 324

Qy    479 ggccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaag 538
Db    325 GGCCGCGAGCAGCTACTGGAGCGCGTTCGCCGCAAGGTACCTG---------CGCTGCGA 375

Qy    539 agtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctg 598
Db    376 GGCGATGACAGTCGATGGCGTCCGGAAGACCTGAGCCGACTGCTGGGAGAGGTGCAAGCT 435

Qy    599 atgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgag 658
Db    436 TTGAGAGGAGTGCAGGAGAGCACGGAGGCACGGCTGCAGGAACTCAGGCAGCAGAACGAG 495

Qy    659 gctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtc 718
Db    496 ATCTTGTGGCGAGAGGTGGTAACTTTGCGGCAGAGCCACAGTCAGCAGCACCGGGTCATC 555

Qy    719 aacaagctcattcagttcctgatctcactggtgcagtca---aaccggatcctgggggtg 775
Db    556 GGCAAGCTAATCCAGTGCCTGTTTGGGCCACTTCAGACAGGGCCCAGCAGTACAGGAGCC 615

Qy    776 aagagaaagatccccctgatgctgaacgacagtggctc 813
Db    616 AAGAGAAAACTGTCCCTAATGCTAGATGAGGGGAGCGC 653
```

RESULT 11
AB029349
LOCUS AB029349 1686 bp mRNA ROD 23-SEP-1999
DEFINITION Mus musculus mRNA for transcription factor HSF4b isoform, complete cds.
ACCESSION AB029349
VERSION AB029349.1 GI:5921136
KEYWORDS transcription factor HSF4b isoform.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Tanabe,M., Sasai,N., Nagata,K., Liu,X.D., Liu,P.C., Thiele,D.J. and Nakai,A.
TITLE The mammalian HSF4 gene generates both an activator and a repressor of heat shock genes by alternative splicing
JOURNAL J. Biol. Chem. 274 (39), 27845-27856 (1999)
MEDLINE 99419073
REFERENCE 2 (bases 1 to 1686)

```
  AUTHORS   Nakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) to the DDBJ/EMBL/GenBank databases. Akira
            Nakai, Institute for Frontier Medical Sciences, Department of
            Molecular and Cell Biology; Sakyo-ku, Kyoto 606-8397, Japan
            (E-mail:nakai@frontier.kyoto-u.ac.jp, Tel:81-75-751-4638,
            Fax:81-75-752-9017)
FEATURES             Location/Qualifiers
     source          1. .1686
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             6. .1484
                     /codon_start=1
                     /product="transcription factor HSF4b isoform"
                     /protein_id="BAA84583.1"
                     /db_xref="GI:5921137"
                     /translation="MQEAPAALPTEPGPSPVPAFLGKLWALVGDPGTDHLIRWSPSGT
                     SFLVSDQSRFAKEVLPQYFKHSNMASFVRQLNMYGFRKVVSIEQGGLLRPERDHVEFQ
                     HPSFVRGREQLLERVRRKVPALRGDDSRWRPEDLSRLLGEVQALRGVQESTEARLQEL
                     RQQNEILWREVVTLRQSHSQQHRVIGKLIQCLFGPLQTGPSSTGAKRKLSLMLDEGSA
                     CSASAKFNACPVSGALLQDPYFIQSPLPETTLGLSPHRARGPIISDIPEDSPSPEGHR
                     LSPSGGCRRVKGLALLKEEPASPGGDGEAGLALAPNECDFCVTAPPPLPVAVVQAILE
                     GKGSYSPEGPRSVQQPEPRGPREVPDRGTLGLDRGNRSPESLLPPMLLRPAPETLEPV
                     APVDVLGPSLHGREWTLMDLDMELSLMQPLAPETDEAELTVKELNSSGVGKDHTLGTP
                     LMLDVQADLEGAALSVPGALTLYNVTESNASYLDPGASPSSP"
BASE COUNT      366 a    521 c    500 g    299 t
ORIGIN


  Query Match             12.0%;  Score 259.6;  DB 12;  Length 1686;
  Best Local Similarity   65.4%;  Pred. No. 1.8e-35;
  Matches  417;  Conservative    0;  Mismatches  209;  Indels   12;  Gaps    2;

Qy     179 cccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtg 238
           |||   | | | || ||||||   || || ||||||||   |||||| ||| | || ||
Db      30 CCCACGGAGCCAGGCCCCAGCCCGGTACCTGCCTTCCTCGGCAAGCTATGGGCGCTGGTA 89

Qy     239 agcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtg 298
           ||||||| | ||||||| |||||| |||||||||||| || | |||||||| |||
Db      90 GGCGACCCAGGCACCGACCACCTCATCCGCTGGAGCCCGAGTGGCACCAGCTTCCTCGTA 149

Qy     299 ttcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatg 358
           || ||| |||  || |||||||| |||||||| |||||||||||||| |||||||
Db     150 AGTGATCAGAGCCGCTTCGCCAAGGAAGTGCTGCCCCAGTACTTCAAGCACAGCAACATG 209

Qy     359 gccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcag 418
           || ||||| || || || |||||||||||||| || ||||| |||||   |||||||||
Db     210 GCGAGCTTTGTTCGCCAACTCAACATGTATGGTTTTCGGAAGGTGGTGAGCATCGAGCAA 269

Qy     419 ggcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgt 478
           || |||||| |||  |||||| | ||| ||   ||||| ||||| ||  ||||| ||||
Db     270 GGTGGCCTGCTCAGACCAGAGCGTGACCACGTTGAGTTTCAGCATCCGAGCTTCGTGCGA 329

Qy     479 ggccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaag 538
           ||||  |||||||| || |||  | |     | || ||  |           | |||
Db     330 GGCCGCGAGCAGCTACTGGAGCGCGTTCGCCGCAAGGTACCTG---------CGCTGCGA 380

Qy     539 agtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctg 598
           | || ||||        || | |||   | | | |   ||||||   || |||||
Db     381 GGCGATGACAGTCGATGGCGTCCGGAAGACCTGAGCCGACTGCTGGGAGAGGTGCAAGCT 440

Qy     599 atgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgag 658
           |||  ||   ||||||| ||| |||  |   ||| | ||   | | |||  |||| |||
Db     441 TTGAGAGGAGTGCAGGAGAGCACGGAGGCACGGCTGCAGGAACTCAGGCAGCAGAACGAG 500

Qy     659 gctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtc 718
           ||||||| |||||||  |   | |||||||  ||    |||| ||    ||| ||
Db     501 ATCTTGTGGCGAGAGGTGGTAACTTTGCGGCAGAGCCACAGTCAGCAGCACCGGGTCATC 560

Qy     719 aacaagctcattcagttcctgatctcactggtgcagtca---aaccggatcctgggggtg 775
           |||||| || |||| |||| |    |   | ||| ||     | | |     || |
Db     561 GGCAAGCTAATCCAGTGCCTGTTTGGGCCACTTCAGACAGGGCCCAGCAGTACAGGAGCC 620

Qy     776 aagagaaagatccccctgatgctgaacgacagtggctc 813
           ||||||||  |  |||| ||||| | ||  |  || |
Db     621 AAGAGAAAACTGTCCCTAATGCTAGATGAGGGGAGCGC 658


RESULT  12
AF059275
LOCUS       AF059275    11395 bp    DNA             ROD       01-DEC-1998
DEFINITION  Mus musculus heat shock transcription factor 1 (Hsf1) gene, partial
            cds.
ACCESSION   AF059275
VERSION     AF059275.1  GI:3126917
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 11395)
  AUTHORS   Zhang,Y., Koushik,S., Dai,R. and Mivechi,N.F.
  TITLE     Structural organization and promoter analysis of murine heat shock
            transcription factor-1 gene
  JOURNAL   J. Biol. Chem. 273 (49), 32514-32521 (1998)
  MEDLINE   99047622
REFERENCE   2  (bases 1 to 11395)
  AUTHORS   Zhang,Y., Koushik,S.V., Dai,R. and Mivechi,N.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1998) Institute of Molecular Medicine & Genetics,
            1120 15th Street, Augusta, GA 30912-2630, USA
FEATURES             Location/Qualifiers
     source          1. .11395
                     /organism="Mus musculus"
                     /strain="129SVJ"
                     /db_xref="taxon:10090"
                     /tissue_type="liver"
                     /sex="female"
                     /dev_stage="4-8 weeks"
                     /clone="3-1"
     intron          <1. .4943
                     /gene="Hsf1"
                     /number=1
     gene            <1. .9886
                     /gene="Hsf1"
     mRNA            join(<4944. .5052,5427. .5563,5810. .5934,6592. .6667,
                     6741. .6802,6925. .7021,7126. .7262,7394. .7663,8918. .9023,
                     9101. .9166,9237. .9306,9387. .9886)
                     /gene="Hsf1"
                     /product="heat shock transcription factor 1"
     CDS             join(<4944. .5052,5427. .5563,5810. .5934,6592. .6667,
                     6741. .6802,6925. .7021,7126. .7262,7394. .7663,8918. .9023,
                     9101. .9166,9237. .9306,9387. .9592)
                     /gene="Hsf1"
                     /codon_start=1
                     /product="heat shock transcription factor 1"
                     /protein_id="AAC80425.1"
                     /db_xref="GI:3126918"
                     /translation="SGNSFHVFDQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVH
                     IEQGGLVKPERDDTEFQHPCFLRGQEQLLENIKRKVTSVSTLKSEDIKIRQDSVTRLL
                     TDVQLMKGKQECMDSKLLAMKHENEALWREVASLRQKHAQQQKVVVNKLIQFLISLVQS
                     NRILGVKRKIPLMLSDSNSAHSVPKYGRQYSLEHVHGPGPYSAPSPAYSSSSLYSSDA
                     VTSSGPIISDITELAPTSPLASPGRSIDERPLSSSTLVRVKQEPPSPPHSPRVLEASP
                     GRPSSMDTPLSPTAFIDSILRESEPTPAASNTAPMDTTGAQAPALPTPSTPEKCLSVA
                     CLDKNELSDHLDAMDSNLDNLQTMLTSHGFSVDTSALLDLFSPSVTMPDMSLPDLDSS
                     LASIQELLSPQEPPRPIEAENSNPDSGKQLVHYTAQPLFLLDPDAVDTGSSELPVLFE
                     LGESSYFSEGDDYTDDPTISLLTGTEPHKAKDPTVS"
     exon            4944. .5052
                     /gene="Hsf1"
                     /number=2
     intron          5053. .5426
                     /gene="Hsf1"
                     /number=2
     exon            5427. .5563
                     /gene="Hsf1"
                     /number=3
```

```
     intron          5564. .5809
                     /gene="Hsf1"
                     /number=3
     exon            5810. .5934
                     /gene="Hsf1"
                     /number=4
     intron          5935. .6591
                     /gene="Hsf1"
                     /number=4
     exon            6592. .6667
                     /gene="Hsf1"
                     /number=5
     intron          6668. .6740
                     /gene="Hsf1"
                     /number=5
     exon            6741. .6802
                     /gene="Hsf1"
                     /number=6
     intron          6803. .6924
                     /gene="Hsf1"
                     /number=6
     exon            6925. .7021
                     /gene="Hsf1"
                     /number=7
     intron          7022. .7125
                     /gene="Hsf1"
                     /number=7
     exon            7126. .7262
                     /gene="Hsf1"
                     /number=8
     intron          7263. .7393
                     /gene="Hsf1"
                     /number=8
     exon            7394. .7663
                     /gene="Hsf1"
                     /number=9
     intron          7664. .8917
                     /gene="Hsf1"
                     /number=9
     exon            8918. .9023
                     /gene="Hsf1"
                     /number=10
     intron          9024. .9100
                     /gene="Hsf1"
                     /number=10
     exon            9101. .9166
                     /gene="Hsf1"
                     /note="isoform"
                     /number=11
     intron          9167. .9236
                     /gene="Hsf1"
                     /number=11
     exon            9237. .9306
                     /gene="Hsf1"
                     /number=12
     intron          9307. .9386
                     /gene="Hsf1"
                     /number=12
     exon            9387. .9886
                     /gene="Hsf1"
                     /number=13
BASE COUNT     2685 a   3044 c   2670 g   2996 t
ORIGIN


  Query Match             11.2%;  Score 241;  DB 12;  Length 11395;
  Best Local Similarity   69.6%;  Pred. No. 2.1e-32;
  Matches  393;  Conservative    0;  Mismatches  150;  Indels   22;  Gaps    4;

Qy   1542 cagggaagcagctggtgcactacacagcgcagccgctgttcctgctggaccccggctccg 1601
          |||| |||||||||||||||||||| || |||||| ||||||||||||||| || |   | |
Db   9384 CAGGAAAGCAGCTGGTGCACTACACGGCTCAGCCTCTGTTCCTGCTGGATCCTGATGCTG 9443

Qy   1602 tggacaccgggagcaacgacctgccggtgctgtttgagctgggagagggctcctacttct 1661
          ||||||| ||||||| || ||||| ||||| |||||||||||| ||| ||||||||||||
Db   9444 TGGACACAGGGAGCAGTGAGCTGCCTGTGCTCTTTGAGCTGGGGGAGAGCTCCTACTTCT 9503

Qy   1662 ccgaaggggacggcttcgccgaggaccccaccatctccctgctgacaggctcggagcctc 1721
          | || ||||| | || | | || || |||||||||||| || ||||||||| | || || |
Db   9504 CTGAGGGGGATGACTACACGGATGATCCCACCATCTCTCTTCTGACAGGCACTGAACCCC 9563

Qy   1722 ccaaagccaaggaccccactgtctcctagaggccccggagg----------agctgggcc 1771
            |||||||||||||||||||||||||||||  | | |  |           ||| |  |
Db   9564 ATAAAGCCAAGGACCCCACTGTCTCCTAGAGCTCTCAGGAGTTGTCAGGCTGGCTTGTGC 9623

Qy   1772 agccgcccaccccccaccccccagtgcagggctggtcttggggaggcagggcagcctcgcgg 1831
               |||   || | ||| ||  || |||||||| | ||||| ||   |||    |  |
Db   9624 CTGGCCCCCAACCTATCCCTAGGACATGGCTGGTCCTAGGGAGACAAAACAGTTGGGTAG 9683

Qy   1832 tcttgggcactggtgggtcggccgccatagccccagtaggaca---aacgggctcgggtc 1888
          ||  ||| ||    ||    | | ||| | |||||||  |  ||   |  |  || ||  |
Db   9684 TCCAGGGGACCCTAGGTCAAGTCACCACAACCCCAGTGGAGCACAGATGGAACTTGGTCC 9743

Qy   1889 tgggcagcacctctggtcaggagggtcaccctg-----gcctgccagtctgccttccccc 1943
          ||||||| ||||  | ||||||||   | ||||      |   |   ||||||| |||
Db   9744 TGGGCAGTACCTTGGATCAGGAGGAAGATCCTGAGGGCTGCATACCTGCTGCCTTTACCC 9803

Qy   1944 aaccccgtgtcctgtggtttggttggggcttcacagccacacctggactgaccctgcagg 2003
           | |||  |  ||    | ||||    |||||||||||||| ||||||||||||||||
Db   9804 CAGCCCCAGGTCTACTCTCTGGTCACAGCTTCACAGCCACACTTGGACTGACCCTGCAGG 9863

Qy   2004 ttgttcatagtcagaattgtattttggatttttacacaactgtcccgttccccgctccac 2063
          ||||||||    | ||||||||||||||||||||  |  ||||| ||   |  || |||
Db   9864 TTGTTCAT----AAAATTGTATTTTGGATTTTTACATGAGAGTCCCCTTTAACCCTTCAC 9919

Qy   2064 agagatacacagatatatacacaca 2088
          | | | | | | ||||| | ||| |
Db   9920 AAATAAATATATATATACATACATA 9944


RESULT  13
CHKHSF3C
LOCUS       CHKHSF3C     2675 bp    mRNA            VRT       17-JUN-1993
DEFINITION  Chicken mRNA sequence.
ACCESSION   L06126
VERSION     L06126.1  GI:289816
KEYWORDS    .
SOURCE      Gallus gallus (sub_species domesticus) (library: lambda ZAPII)
            embryo blood cDNA to mRNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2675)
  AUTHORS   Nakai,A. and Morimoto,R.I.
  TITLE     Characterization of a novel chicken heat sock transcription factor,
            heat shock factor 3, suggests a new regulatory pathway
  JOURNAL   Mol. Cell. Biol. 13, 1983-1997 (1993)
  MEDLINE   93204945
FEATURES             Location/Qualifiers
     source          1. .2675
                     /organism="Gallus gallus"
                     /sub_species="domesticus"
                     /db_xref="taxon:9031"
                     /cell_type="red blood cell"
                     /dev_stage="embryo"
                     /tissue_type="blood"
                     /tissue_lib="lambda ZAPII"
BASE COUNT      731 a    587 c    651 g    706 t
ORIGIN


  Query Match             10.7%;  Score 230.4;  DB 4;  Length 2675;
  Best Local Similarity   61.2%;  Pred. No. 1.7e-30;
  Matches  394;  Conservative    0;  Mismatches  241;  Indels    9;  Gaps    1;

Qy    151 cttgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggc 210
```

```
Db      4 CCTGAGCATGAGGGAAGGGTCGGCGCTGCCCGGCGCCCCCGGCGCCGCCCCGGTGCCTGG 63

Qy    211 cttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctg 270
Db     64 CTTCCTGGCCAAGCTGTGGGCCCTGGTGGAGGATCCGCAGAGCGACGACGTCATCTGCTG 123

Qy    271 gagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgct 330
Db    124 GAGCCGGAATGGCGAGAACTTCTGCATCCTGGATGAGCAGCGGTTTGCCAAGGAGCTGCT 183

Qy    331 gcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtatgg 390
Db    184 CCCCAAGTACTTCAAACACAACAACATCTCCAGCTTCATCCGACAGCTCAACATGTATGG 243

Qy    391 cttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgacac 450
Db    244 TTTCAGGAAGGTGGTTGCTTTGGAGAATGGTATGATTACAGCAGAGAAAAACTCAGTCAT 303

Qy    451 ggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaagag 510
Db    304 TGAGTTCCAGCACCCTTTCTTCAAGCAAGGAAATGCACATTTATTGGAAAACATCAAACG 363

Qy    511 gaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgt 570
Db    364 ---------CAAGGTGTCTGCAGTAAGAACTGAGGATCTCAAGGTCTGTGCAGAGGACTT 414

Qy    571 caccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaa 630
Db    415 ACACAAAGTACTCTCTGAAGTCCAGGAAATGAGAGAGCAGCAGAATAACATGGATATCAG 474

Qy    631 gctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagccttcggca 690
Db    475 ACTGGCTAACATGAAGAGAGAAAATAAGGCCCTGTGGAAAGAAGTGGCAGTTCTAAGGCA 534

Qy    691 gaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggt 750
Db    535 GAAGCACAGTCAACAACAGAAGCTGCTGTCTAAGATTCTTCAGTTTATACTAAGTTTGAT 594

Qy    751 gcagtcaaaccggatcctgggggtgaagagaaagatccccctga 794
Db    595 GCGAGGAAATTATATTGTTGGGGTCAAAAGGAAAAGGTCTCTGA 638


RESULT  14
MMHSF2
LOCUS       MMHSF2       1972 bp    mRNA            ROD       28-NOV-1991
DEFINITION  M.musculus mRNA for heat shock transcripton factor 2.
ACCESSION   X61754
VERSION     X61754.1  GI:51447
KEYWORDS    heat shock transcription factor; HSF2.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1972)
  AUTHORS   Sarge,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1991) K.D. Sarge, Northwestern University, Dep of
            Biochem.,Mol. and Cell. Biology, Hogan 2-100, 2153 Sheridan Rd.,
            Evanston IL 60208, USA
REFERENCE   2  (bases 1 to 1972)
  AUTHORS   Sarge,K.D., Zimarino,V., Holm,K., Wu,C. and Morimoto,R.I.
  TITLE     Cloning and characterization of two mouse heat shock factors with
            distinct inducible and constitutive DNA-binding ability
  JOURNAL   Genes Dev. 5 (10), 1902-1911 (1991)
  MEDLINE   92009180
COMMENT     See also x61753.
FEATURES             Location/Qualifiers
     source          1. .1972
                     /organism="Mus musculus"
                     /strain="Balb/C"
                     /db_xref="taxon:10090"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="Clontech lambda gt11 liver cDNA"
                     /clone="C9"
     CDS             159.. .1712
                     /codon_start=1
                     /product="heat shock transcription factor 2"
                     /protein_id="CAA43893.1"
                     /db_xref="GI:51448"
                     /db_xref="MGD:MGI:96239"
                     /db_xref="SWISS-PROT:P38533"
                     /translation="MKQSSNVPAFLSKLWTLVEETHTNEFITWSQNGQSFLVLDEQRF
                     AKEILPKYFKHNNMASFVRQLNMYGFRKVVHIESGIIKQERDGPVEFQHPYFKQGQDD
                     LLENIKRKVSSSKPEENKIRQEDLTKIISSAQKVQIKQETIESRLSELKSENESLWKE
                     VSELRAKHAQQQQVIRKIVQFIVTLVQNNQLVSLKRKRPLLLNTNGAPKKNLYQHIVK
                     EPTDNHHHKVPHSRTEGLKSRERISDDIIIYDVTDDNVDEENIPVIPETNEDVVVDSS
                     NQYPDIVIVEDDNEDEYAPVIQSGEQSEPAREPLRVGSAGSSSPLMSSAVQLNGSSSL
                     TSEDPVTMMDSILNDNINLLGKVELLDYLDSIDCSLEDFQAMLSGRQFSIDPDLLVDS
                     ENKGLEATKSSVVQHVSEEGRKSKSKPDKQLIQYTAFPLLAFLDGNSASAIEQGSTTA
                     SSEVVPSVDKPIEVDELLDSSLDPEPTQSKLVRLEPLTEAEASEATLFYLCELAPAPL
                     DSDMPLLDS"
BASE COUNT      604 a    420 c    432 g    516 t
ORIGIN


  Query Match             9.5%;  Score 204.2;  DB 12;  Length 1972;
  Best Local Similarity   59.4%;  Pred. No. 5.4e-26;
  Matches  369;  Conservative    0;  Mismatches  243;  Indels    9;  Gaps    1;

Qy    199 caacgtcccggccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgc 258
Db    173 CAACGTGCCGGCTTTCCTCAGCAAGCTGTGGACGCTTGTGGAGGAAACCCACACCAACGA 232

Qy    259 gctcatctgctggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgc 318
Db    233 GTTCATCACCTGGAGTCAGAATGGACAAAGTTTTCTGGTCTTGGATGAGCAAAGATTTGC 292

Qy    319 caaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagct 378
Db    293 AAAGGAAATTCTTCCTAAGTACTTCAAACACAATAACATGGCGAGCTTTGTGAGACAACT 352

Qy    379 caacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccaga 438
Db    353 AAATATGTATGGCTTCCGAAAAGTAGTGCATATCGAATCTGGAATTATCAAACAGGAAAG 412

Qy    439 gagagacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttga 498
Db    413 AGATGGCCCTGTTGAATTTCAGCATCCTTATTTCAAGCAAGGCCAGGATGACCTGTTGGA 472

Qy    499 gaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccg 558
Db    473 GAACATTAAAAGGAAG---------GTTTCATCTTCAAAACCAGAGGAAAATAAAATTCG 523

Qy    559 ccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtg 618
Db    524 TCAGGAAGATTTAACAAAAATTATTAGTAGTGCTCAGAAGGTTCAAATAAAACAAGAAAC 583

Qy    619 catggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggc 678
Db    584 TATTGAGTCCAGGCTTTCAGAATTAAAAAGTGAGAATGAATCCCTTTGGAAGGAGGTGTC 643

Qy    679 cagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcct 738
Db    644 AGAACTAAGAGCAAAGCATGCCCAGCAGCAACAAGTTATTCGGAAGATTGTCCAGTTTAT 703

Qy    739 gatctcactggtgcagtcaaaccggatcctgggggtgaagagaaagatccccctgatgct 798
Db    704 TGTTACATTGGTTCAGAATAATCAACTTGTGAGTTTAAAACGTAAAAGGCCTCTACTTCT 763

Qy    799 gaacgacagtggctcagcaca 819
Db    764 AAACACAAATGGAGCCCCAAA 784
```

```
RESULT 15
CHKHSF3B
LOCUS       CHKHSF3B     2366 bp    mRNA            VRT       17-JUN-1993
DEFINITION  Chicken mRNA sequence.
ACCESSION   L06125
VERSION     L06125.1  GI:289815
KEYWORDS    .
SOURCE      Gallus gallus (sub_species domesticus) cDNA to mRNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2366)
  AUTHORS   Nakai,A. and Morimoto,R.I.
  TITLE     Characterization of a novel chicken heat sock transcription factor,
            heat shock factor 3, suggests a new regulatory pathway
  JOURNAL   Mol. Cell. Biol. 13, 1983-1997 (1993)
  MEDLINE   93204945
FEATURES             Location/Qualifiers
     source          1. .2366
                     /organism="Gallus gallus"
                     /sub_species="domesticus"
                     /db_xref="taxon:9031"
BASE COUNT      719 a    498 c    531 g    618 t
ORIGIN


  Query Match             9.4%;  Score 203.6;  DB 4;  Length 2366;
  Best Local Similarity   59.3%;  Pred. No. 6.6e-26;
  Matches  369;  Conservative    0;  Mismatches  244;  Indels    9;  Gaps    1;

Qy    203 gtcccggccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctc 262
          || |||||||||| | |||||||||| | || ||| |||| || || | ||  |||
Db     74 GTGCCGGCCTTCCTCAGCAAGCTGTGGGCGCTGGTGGGCGAGGCGCCCAGCAACCAGCTT 133

Qy    263 atctgctggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaag 322
          |||  |||||||| ||  ||  | || |||   ||||| ||  |       ||||| ||
Db    134 ATCACCTGGAGCCAGAATGGCCAGAGTTTCTTGGTGTTGGATGAACAGAGATTTGCAAAA 193

Qy    323 gaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaac 382
          ||| | ||||| |||||||||||||||||||||||||||| ||||| ||  | ||||| |||
Db    194 GAGATTCTGCCTAAGTACTTCAAGCACAACAACATGGCAAGCTTTGTCAGACAGCTAAAC 253

Qy    383 atgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagaga 442
          |||||||||||||| ||||| |||||  | ||    ||      |       |  |
Db    254 ATGTATGGCTTCCGTAAAGTTGTCCATGTTGACTCTGGGATTGTCAAACTGGAACGAGAT 313

Qy    443 gacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaac 502
          |        ||||| |||||||| |  ||   ||  || |  ||| |  |  | ||  ||
Db    314 GGTCTAGTTGAGTTTCAGCACCCGTATTTTAAGCAGGGTCGAGAGGACTTGTTGGAACAC 373

Qy    503 atcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccag 562
          || || ||||| ||          ||  |    |    |||||| |  |||||  |||||
Db    374 ATTAAAAGGAAGGTT---------TCTTCTTCAAGACCTGAAGAAAACAAGATAAGCCAG 424

Qy    563 gacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatg 622
          ||      || ||||  |  | |      |  |||    || ||   || || |||   ||
Db    425 GAGGATCTCTCCAAAATAATAAGTAGTGCTCAGAAAGTGGAGATTAAACAAGAGACTATT 484

Qy    623 gactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagc 682
          || ||   | |    ||  ||||    |||||||  |||| ||| |||| |||||
Db    485 GAATCTCGGTTGTCTGCTCTGAAAAGGGAGAATGAATCTCTTTGGAGGGAAGTGGCAGAA 544

Qy    683 cttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatc 742
          ||  |    || ||     | |||||| ||||  |    ||| |  | |||||  |  |
Db    545 CTGAGAGCAAAACACCTGAAACAACAGCAAGTTATTCGGAAGATTGTACAGTTTATTGTT 604

Qy    743 tcactggtgcagtcaaaccggatcctgggggtgaagagaaagatccccctgatgctgaac 802
           |  ||||||||   ||||   |  || |  | ||  | ||||  || ||  | ||||||
Db    605 ACCTTGGTGCAGAATAACCAACTAGTGAGCCTAAAACGCAAGAGGCCTCTACTTCTGAAC 664

Qy    803 gacagtggctcagcacattcca 824
          | |||  |  || | || |
Db    665 ACTAATGGACCTACAAAGTCGA 686


Search completed: March  6, 2000, 23:49:41
Job time: 3589 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2000, 20:18:14 ; Search time 962.55 Seconds
(without alignments)
98.064 Million cell updates/sec

Title: US-09-304-121-3
Perfect score: 25
Sequence: 1 ngaannttcnnnnnnnttcnngaan 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	48.0	94	40	D43321	D43321 D43321 Rice
c	2	12	48.0	120	28	C12071	C12071 C12071 Yuji

	3	12	48.0	136	20	D25791 D25791 HUMGS04159
	4	12	48.0	169	81	B81977 B81977 RPCI11-19M2
c	5	12	48.0	177	62	AI873729 AI873729 wm29c09.x
	6	12	48.0	178	48	AI595065 AI595065 vb76e06.y
	7	12	48.0	184	39	AA905401 AA905401 oj84g11.s
	8	12	48.0	188	59	AV114853 AV114853
	9	12	48.0	195	70	AV230532 AV230532 AV230532
c	10	12	48.0	197	70	AV237498 AV237498 AV237498
	11	12	48.0	207	30	AA271613 AA271613 vb76e06.r
	12	12	48.0	208	51	AU068731 AU068731 AU068731
	13	12	48.0	209	63	AI933599 AI933599 wn41c09.x
	14	12	48.0	211	27	C08244 C08244 C08244 Yuji
	15	12	48.0	211	28	AA071319 AA071319 zm73h10.r
c	16	12	48.0	213	26	W44843 W44843 zc78g08.r1
	17	12	48.0	213	61	AI827151 AI827151 wl08h04.x
	18	12	48.0	218	70	AV248478 AV248478 AV248478
	19	12	48.0	221	71	AV283090 AV283090 AV283090
	20	12	48.0	227	64	AW035966 AW035966 EST282825
c	21	12	48.0	228	64	AW079585 AW079585 xc19c10.x
	22	12	48.0	231	70	AV252321 AV252321 AV252321
	23	12	48.0	233	50	AV043755 AV043755 AV043755
	24	12	48.0	233	62	AI888421 AI888421 wn20g10.x
	25	12	48.0	234	71	AV279890 AV279890 AV279890
c	26	12	48.0	235	73	AV335705 AV335705 AV335705
c	27	12	48.0	236	72	AV292990 AV292990 AV292990
	28	12	48.0	237	45	AI376570 AI376570 te64g06.x
	29	12	48.0	238	51	AV048190 AV048190 AV048190
	30	12	48.0	241	62	AI888609 AI888609 wn33g01.x
c	31	12	48.0	241	74	AV380482 AV380482 AV380482
c	32	12	48.0	243	70	AV248858 AV248858 AV248858
	33	12	48.0	245	62	AI925329 AI925329 wn45e07.x
	34	12	48.0	246	71	AV254627 AV254627 AV254627
	35	12	48.0	246	71	AV281906 AV281906 AV281906
	36	12	48.0	247	71	AV320489 AV320489 AV320489
	37	12	48.0	249	71	AV257886 AV257886 AV257886
	38	12	48.0	250	70	AV247480 AV247480 AV247480
	39	12	48.0	252	91	AQ106130 AQ106130 HS_3055_A
	40	12	48.0	254	71	AV270362 AV270362 AV270362
	41	12	48.0	256	28	C17160 C17160 C17160 Clon
	42	12	48.0	257	20	T01814 T01814 wEST02535 E
	43	12	48.0	258	20	Z18218 Z18218 ATTS0723 Gr
	44	12	48.0	258	32	AA376612 AA376612 EST89057
	45	12	48.0	258	71	AV268419 AV268419 AV268419

ALIGNMENTS

```
RESULT   1
D43321
LOCUS       D43321         94 bp    mRNA            EST       04-MAY-1998
DEFINITION  D43321 Rice callus cDNA (H.Uchimiya) Oryza sativa cDNA clone SS249,
            mRNA sequence.
ACCESSION   D43321
VERSION     D43321.1  GI:3107581
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 94)
  AUTHORS   Uchimiya,H.
  TITLE     On nucleotide sequence of Oryza sativa
  JOURNAL   Unpublished (1994)
COMMENT     On May 8, 1995 this sequence version replaced gi:801268.
            Contact: Hirofumi Uchimiya
            Institute of Mol. & Cell. Bioscience, Department of Cellular
            Function
            The University of Tokyo
            1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan
            Tel: 03-3812-2111(ex.7844)
            Fax: 03-3812-2910
            Email: huchimiy@tansei.cc.u-tokyo.ac.jp
            PROJECT= Uchimiya.
FEATURES             Location/Qualifiers
     source          1. .94
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /clone="SS249"
                     /clone_lib="Rice callus cDNA (H.Uchimiya)"
                     /tissue_type="callus"
BASE COUNT       15 a     28 c     26 g     17 t      8 others
ORIGIN


  Query Match           48.0%;  Score 12;  DB 40;  Length 94;
  Best Local Similarity   52.2%;  Pred. No. 2.7e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db      2 GAAGCTTCTGGAAGGTTCTCGAA 24


RESULT   2
C12071/c
LOCUS       C12071         120 bp    mRNA            EST       28-DEC-1998
DEFINITION  C12071 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk145e3 5', mRNA sequence.
ACCESSION   C12071
VERSION     C12071.1  GI:1559624
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 120)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1282252.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .120
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk145e3"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       51 a     15 c     20 g     31 t      3 others
ORIGIN


  Query Match           48.0%;  Score 12;  DB 28;  Length 120;
  Best Local Similarity   52.2%;  Pred. No. 2.7e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     80 GAATTTTCCGGTGTTTTCTAGAA 58


RESULT   3
D25791
LOCUS       D25791         136 bp    mRNA            EST       30-NOV-1995
```

```
DEFINITION  HUMGS04159 Human colon mucosa Homo sapiens cDNA clone cm0533 3',
            mRNA sequence.
ACCESSION   D25791
VERSION     D25791.1  GI:500474
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136)
  AUTHORS   Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.
  TITLE     Global analysis of gene expression in colon mucosa: a large scale
            random cDNA sequencing analysis
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES             Location/Qualifiers
     source          1. .136
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="cm0533"
                     /clone_lib="Human colon mucosa"
                     /note="Adult male, tissue_type = colon mucosa "
BASE COUNT       43 a     22 c     29 g     42 t
ORIGIN
```

  Query Match 48.0%; Score 12; DB 20; Length 136;
  Best Local Similarity 52.2%; Pred. No. 2.8e+03;
  Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||  |||
Db      43 GAAGATTCCTGGTATTTCTCGAA 65
```

```
RESULT   4
B81977
LOCUS       B81977         169 bp    DNA             GSS       09-APR-1999
DEFINITION  RPCI11-19M2.TP RPCI-11 Homo sapiens genomic clone RPCI-11-19M2,
            genomic survey sequence.
ACCESSION   B81977
VERSION     B81977.1  GI:2869000
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .169
                     /organism="Homo sapiens"
                     /db_xref="GDB:7507201"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-19M2"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       47 a     22 c     43 g     57 t
ORIGIN
```

  Query Match 48.0%; Score 12; DB 81; Length 169;
  Best Local Similarity 52.2%; Pred. No. 2.8e+03;
  Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||  |||
Db      26 GAAACTTCACTCATTTTCTGGAA 48
```

```
RESULT   5
AI873729/c
LOCUS       AI873729       177 bp    mRNA            EST       01-SEP-1999
DEFINITION  wm29c09.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2437360 3'
            similar to SW:GUAA_HUMAN P49915 GMP SYNTHASE
            [GLUTAMINE-HYDROLYZING] ;, mRNA sequence.
ACCESSION   AI873729
VERSION     AI873729.1  GI:5547778
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 1, 1997 this sequence version replaced gi:2059584.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .177
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2437360"
                     /clone_lib="NCI_CGAP_Ut4"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT       53 a     38 c     33 g     53 t
ORIGIN
```

  Query Match 48.0%; Score 12; DB 62; Length 177;
  Best Local Similarity 52.2%; Pred. No. 2.8e+03;

```
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     92 GAAGATTCCTGGTATTTCTCGAA 70


RESULT   6
AI595065
LOCUS       AI595065      178 bp    mRNA            EST       21-APR-1999
DEFINITION  vb76e06.y1 Soares mouse 3NME12 5 Mus musculus cDNA clone
            IMAGE:762946 5' similar to TR:Q60441 Q60441 SUPPRESSOR OF LEC15
            GLYCOSYLATION MUTATION SL15. ;, mRNA sequence.
ACCESSION   AI595065
VERSION     AI595065.1  GI:4604113
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 178)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3186981.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:463866
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .178
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:762946"
                     /clone_lib="Soares mouse 3NME12 5"
                     /sex="unknown"
                     /tissue_type="fetus"
                     /dev_stage="12.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTTT
                     3'],  on total  mouse RNA [provided by Minoru Ko, Wayne
                     State Univ.]; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT       31 a     54 c     39 g     54 t
ORIGIN


  Query Match            48.0%;  Score 12;  DB 48;  Length 178;
  Best Local Similarity  52.2%;  Pred. No. 2.8e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     75 GAATCTTCACTTCTGTTCAGGAA 97


RESULT   7
AA905401
LOCUS       AA905401      184 bp    mRNA            EST       09-JUN-1998
DEFINITION  oj84g11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1505060 3', mRNA sequence.
ACCESSION   AA905401
VERSION     AA905401.1  GI:3040524
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:2754419.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1372   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 161.
FEATURES             Location/Qualifiers
     source          1. .184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1505060"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       82 a     25 c     22 g     55 t
ORIGIN


  Query Match            48.0%;  Score 12;  DB 39;  Length 184;
  Best Local Similarity  52.2%;  Pred. No. 2.8e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     16 GAAGATTCAAAAATTTTCCTGAA 38


RESULT   8
AV114853
LOCUS       AV114853      188 bp    mRNA            EST       30-JUN-1999
DEFINITION  AV114853 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA
            clone 2610037H24, mRNA sequence.
ACCESSION   AV114853
VERSION     AV114853.1  GI:5297004
KEYWORDS    EST.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 188)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1134722.
            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .188
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="2610037H24"
                      /clone_lib="Mus musculus C57BL/6J 10-day embryo"
                      /sex="mixed"
                      /dev_stage="10-day embryo"
BASE COUNT       55 a     36 c     43 g     54 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 59;  Length 188;
  Best Local Similarity   52.2%;  Pred. No. 2.8e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    102 GAAGATTCCTGGAATTTCTAGAA 124


RESULT  9
AV230532
LOCUS       AV230532      195 bp    mRNA            EST       03-NOV-1999
DEFINITION  AV230532 RIKEN full-length enriched, 0 day neonate skin Mus
            musculus cDNA clone 4631432O17 3', mRNA sequence.
ACCESSION   AV230532
VERSION     AV230532.1  GI:6183047
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 195)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5405906.
            Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .195
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="4631432O17"
                      /clone_lib="RIKEN full-length enriched, 0 day neonate
                      skin"
                      /sex="mixed"
                      /tissue_type="skin"
                      /dev_stage="0 day neonate"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 100.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                      3']. cDNA was cloned into the XhoI and BamHI sites.
                      Vector: a modified pBluescript KS(+) after bulk excision
                      from Lambda FLC I"
BASE COUNT       69 a     40 c     34 g     52 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 70;  Length 195;
  Best Local Similarity   52.2%;  Pred. No. 2.9e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     82 GAAAATTCGCTGCTTTTCAGGAA 104


RESULT  10
AV237498/c
```

```
LOCUS       AV237498      197 bp    mRNA            EST       03-NOV-1999
DEFINITION  AV237498 RIKEN full-length enriched, 10 day neonate skin Mus
            musculus cDNA clone 4732421H07 3' similar to AF061026 Mus musculus
            leucine zipper-EF-hand containing transmembrane protein 1 (Letm1)
            mRNA, mRNA sequence.
ACCESSION   AV237498
VERSION     AV237498.1  GI:6190010
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189134.
            Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .197
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4732421H07"
                     /clone_lib="RIKEN full-length enriched, 10 day neonate
                     skin"
                     /sex="mixed"
                     /tissue_type="skin"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I"
BASE COUNT       65 a     41 c     42 g     49 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 70;  Length 197;
  Best Local Similarity   52.2%;  Pred. No. 2.9e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||  |||
Db     193 GAAGTTTCTGATTTATTCCAGAA 171


RESULT  11
AA271613
LOCUS       AA271613      207 bp    mRNA            EST       26-MAR-1997
DEFINITION  vb76e06.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone
            IMAGE:762946 5' similar to TR:G1323704 G1323704 SL15. ;, mRNA
            sequence.
ACCESSION   AA271613
VERSION     AA271613.1  GI:1910203
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1402221.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:463866
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 152.
FEATURES             Location/Qualifiers
     source          1. .207
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:762946"
                     /clone_lib="Soares mouse 3NME12 5"
                     /sex="unknown"
                     /tissue_type="fetus"
                     /dev_stage="12.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTTT
                     3'],  on total  mouse RNA [provided by Minoru Ko, Wayne
                     State Univ.]; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
```

Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 42 a 63 c 47 g 55 t
ORIGIN

Query Match 48.0%; Score 12; DB 30; Length 207;
Best Local Similarity 52.2%; Pred. No. 2.9e+03;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 gaannttcnnnnnnnttcnngaa 24
||| ||| ||| |||
Db 76 GAATCTTCACTTCTGTTCAGGAA 98

RESULT 12
AU068731
LOCUS AU068731 208 bp mRNA EST 07-JUN-1999
DEFINITION AU068731 Rice callus Oryza sativa cDNA clone C50295_1A, mRNA sequence.
ACCESSION AU068731
VERSION AU068731.1 GI:5003582
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 208)
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from callus (1998)
JOURNAL Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187046.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .208
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C50295_1A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid. "
BASE COUNT 52 a 57 c 59 g 28 t 12 others
ORIGIN

Query Match 48.0%; Score 12; DB 51; Length 208;
Best Local Similarity 56.5%; Pred. No. 2.9e+03;
Matches 13; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 gaannttcnnnnnnnttcnngaa 24
||| ||| |||| |||
Db 70 GAAGCTTCGAGAATNTTCCGGAA 92

RESULT 13
AI933599
LOCUS AI933599 209 bp mRNA EST 02-SEP-1999
DEFINITION wn41c09.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2448016 3' similar to gb:Z11887 MATRILYSIN PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION AI933599
VERSION AI933599.1 GI:5672936
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 209)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135223.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 110.
FEATURES Location/Qualifiers
source 1. .209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2448016"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"
BASE COUNT 57 a 29 c 44 g 79 t
ORIGIN

Query Match 48.0%; Score 12; DB 63; Length 209;
Best Local Similarity 52.2%; Pred. No. 2.9e+03;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 gaannttcnnnnnnnttcnngaa 24
||| ||| ||| |||
Db 90 GAAGTTTCTATTTCTTTCTTGAA 112

RESULT 14
C08244
LOCUS C08244 211 bp mRNA EST 18-OCT-1999
DEFINITION C08244 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo Caenorhabditis elegans cDNA clone yk184a12 3', mRNA sequence.
ACCESSION C08244
VERSION C08244.1 GI:1533315
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 211)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On May 18, 1995 this sequence version replaced gi:811127.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics

```
           Yata 1111, Mishima, Shizuoka 411, Japan
           Tel: 81-559-81-6854
           Fax: 81-559-81-6855
           Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .211
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk184a12"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       78 a     15 c     42 g     76 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 27;  Length 211;
  Best Local Similarity   52.2%;  Pred. No. 2.9e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||  |||
Db      66 GAATTTTCAAAGAAATTCTGGAA 88


RESULT  15
AA071319
LOCUS       AA071319      211 bp    mRNA            EST       23-DEC-1997
DEFINITION  zm73h10.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
            clone IMAGE:531331 5', mRNA sequence.
ACCESSION   AA071319
VERSION     AA071319.1  GI:1578681
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1282353.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 966   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 180.
FEATURES             Location/Qualifiers
     source          1. .211
                     /organism="Homo sapiens"
                     /db_xref="GDB:3920843"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:531331"
                     /clone_lib="Stratagene neuroepithelium (#937231)"
                     /dev_stage="Ntera-2/RA neuroepithelial cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. NT2
                     cells (Ntera-2/c1.D1) induced with Retinoic Acid for 24
                     hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       59 a     37 c     43 g     72 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 28;  Length 211;
  Best Local Similarity   52.2%;  Pred. No. 2.9e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||  |||
Db      30 GAATTTTCACTGTGTTTCATGAA 52


Search completed: March  6, 2000, 20:18:17
Job time: 1094 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2000, 00:19:53 ; Search time 45.51 Seconds
(without alignments)
65.767 Million cell updates/sec

Title: US-09-304-121-3
Perfect score: 25
Sequence: 1 ngaannttcnnnnnnnttcnngaan 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	48.0	152	2	US-08-463-660-1	Sequence 1, Appli
	2	12	48.0	152	2	US-08-678-280-1	Sequence 1, Appli
	3	12	48.0	584	3	US-08-937-540-12	Sequence 12, Appl
c	4	12	48.0	2022	3	US-08-937-540-7	Sequence 7, Appli
	5	12	48.0	2237	2	US-08-463-620-1	Sequence 1, Appli
	6	12	48.0	2237	4	US-08-224-917-1	Sequence 1, Appli
	7	12	48.0	2237	4	US-08-914-853-1	Sequence 1, Appli
	8	12	48.0	2237	6	PCT-US95-03934A-1	Sequence 1, Appli
c	9	12	48.0	2628	4	US-08-696-944-1	Sequence 1, Appli
c	10	11	44.0	88	1	US-08-433-126A-68	Sequence 68, Appl
c	11	11	44.0	88	2	US-08-433-124A-68	Sequence 68, Appl
c	12	11	44.0	88	6	PCT-US96-06059-68	Sequence 68, Appl
c	13	11	44.0	371	5	US-08-659-188-24	Sequence 24, Appl
c	14	11	44.0	394	2	US-08-650-275-12	Sequence 12, Appl
c	15	11	44.0	394	5	US-09-181-318-12	Sequence 12, Appl
c	16	11	44.0	654	3	US-08-468-819-73	Sequence 73, Appl
c	17	11	44.0	654	3	US-08-468-819-75	Sequence 75, Appl
c	18	11	44.0	800	4	US-08-929-302-3	Sequence 3, Appli
c	19	11	44.0	800	4	US-09-038-014-3	Sequence 3, Appli
c	20	11	44.0	837	2	US-08-371-082-1	Sequence 1, Appli
	21	11	44.0	882	2	US-08-628-291-3	Sequence 3, Appli
	22	11	44.0	882	4	US-09-128-722-3	Sequence 3, Appli
	23	11	44.0	889	2	US-08-832-883-52	Sequence 52, Appl
	24	11	44.0	889	3	US-08-832-877-52	Sequence 52, Appl
	25	11	44.0	999	2	US-08-469-649-1	Sequence 1, Appli
	26	11	44.0	1065	4	US-08-512-955-1	Sequence 1, Appli
c	27	11	44.0	1078	3	US-08-555-723B-1	Sequence 1, Appli
c	28	11	44.0	1079	1	US-08-270-583-1	Sequence 1, Appli
c	29	11	44.0	1079	2	US-08-783-889A-1	Sequence 1, Appli
	30	11	44.0	1204	2	US-08-628-291-11	Sequence 11, Appl
	31	11	44.0	1204	4	US-09-128-722-11	Sequence 11, Appl
c	32	11	44.0	1756	7	5281520-4	Patent No. 5281520
	33	11	44.0	1856	4	US-08-360-606B-29	Sequence 29, Appl
	34	11	44.0	1944	4	US-08-844-056-1	Sequence 1, Appli
c	35	11	44.0	2022	6	PCT-US96-00996-4	Sequence 4, Appli
c	36	11	44.0	2027	4	US-08-377-309-1	Sequence 1, Appli
c	37	11	44.0	2163	7	5281520-1	Patent No. 5281520
c	38	11	44.0	2163	7	5281520-2	Patent No. 5281520
c	39	11	44.0	2174	3	US-08-665-040-1	Sequence 1, Appli
c	40	11	44.0	2291	7	5281520-3	Patent No. 5281520
	41	11	44.0	2339	1	US-08-258-639A-1	Sequence 1, Appli
	42	11	44.0	2339	4	US-08-900-951-1	Sequence 1, Appli
	43	11	44.0	2339	6	PCT-US95-07391A-1	Sequence 1, Appli
c	44	11	44.0	2504	1	US-08-121-713D-63	Sequence 63, Appl
c	45	11	44.0	2504	2	US-08-835-268-63	Sequence 63, Appl

ALIGNMENTS

```
RESULT   1
US-08-463-660-1
; Sequence 1, Application US/08463660
; Patent No. 5759776
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, HELENE S.
;    APPLICANT:  CHEN, LING-CHUN
;    TITLE OF INVENTION:  TARGETS FOR BREAST CANCER DIAGNOSIS AND TREATMENT
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,660
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CIOTTI, THOMAS E.
;      REGISTRATION NUMBER:  21,013
;      REFERENCE/DOCKET NUMBER:  28888-20001.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  152 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..152
US-08-463-660-1


  Query Match             48.0%;  Score 12;  DB 2;  Length 152;
  Best Local Similarity   52.2%;  Pred. No. 42;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;
```

```
Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     47 GAAACTTCAGAATACTTCAAGAA 69


RESULT   2
US-08-678-280-1
; Sequence 1, Application US/08678280
; Patent No. 5776683
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, HELENE S.
;    APPLICANT:  CHUN, LING-CHEN
;    TITLE OF INVENTION:  TARGETS FOR BREAST CANCER DIAGNOSIS AND
;    TITLE OF INVENTION:  TREATMENT
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/678,280
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schiff, J. Michael
;      REGISTRATION NUMBER:  40,253
;      REFERENCE/DOCKET NUMBER:  28888-20001.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  152 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..152
US-08-678-280-1


  Query Match             48.0%;  Score 12;  DB 2;  Length 152;
  Best Local Similarity   52.2%;  Pred. No. 42;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     47 GAAACTTCAGAATACTTCAAGAA 69


RESULT   3
US-08-937-540-12
; Sequence 12, Application US/08937540
; Patent No. 5891697
;  GENERAL INFORMATION:
;    APPLICANT:  Croteau, Rodney B
;    APPLICANT:  Wise, Mitchell L
;    APPLICANT:  Savage, Thomas J
;    APPLICANT:  Katahira, Eva J
;    TITLE OF INVENTION:  Monoterpene Synthases from Common Sage
;    TITLE OF INVENTION:  (Salvia officinalis)
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON & KINDNESS
;      STREET:  1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/937,540
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  WSUR111254
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206 695 1718
;      TELEFAX:  206 224 0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  584 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Salvia officinalis
;    IMMEDIATE SOURCE:
;      CLONE:  Low affinity cDNA probe
US-08-937-540-12


  Query Match             48.0%;  Score 12;  DB 3;  Length 584;
  Best Local Similarity   52.2%;  Pred. No. 52;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    236 GAAGCTTCTCCACAATTCCTGAA 258


RESULT   4
US-08-937-540-7/c
; Sequence 7, Application US/08937540
; Patent No. 5891697
;  GENERAL INFORMATION:
;    APPLICANT:  Croteau, Rodney B
;    APPLICANT:  Wise, Mitchell L
;    APPLICANT:  Savage, Thomas J
;    APPLICANT:  Katahira, Eva J
;    TITLE OF INVENTION:  Monoterpene Synthases from Common Sage
;    TITLE OF INVENTION:  (Salvia officinalis)
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON & KINDNESS
;      STREET:  1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/937,540
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Shelton, Dennis K
;       REGISTRATION NUMBER:  26,997
;       REFERENCE/DOCKET NUMBER:  WSUR111254
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206 695 1718
;       TELEFAX:  206 224 0779
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2022 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Salvia officinalis
;     IMMEDIATE SOURCE:
;       CLONE:  Unknown monoterpene synthase-like sequence
US-08-937-540-7


  Query Match             48.0%;  Score 12;  DB 3;  Length 2022;
  Best Local Similarity   52.2%;  Pred. No. 64;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    999 GAAGCTTCTCCACAATTCCTGAA 977


RESULT   5
US-08-463-620-1
; Sequence 1, Application US/08463620
; Patent No. 5789216
;  GENERAL INFORMATION:
;    APPLICANT:  Lou, Lillian Lien-Li
;    APPLICANT:  Barnett, Jimmy Wayne
;    TITLE OF INVENTION:  Cloning and Expression of Human GMP
;    TITLE OF INVENTION:  Synthetase, its use in Screening for Inhibitors of
;    TITLE OF INVENTION:  Human
;    TITLE OF INVENTION:  GMP Synthetase and Inhibitors of Human GMP Synthetase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Syntex (USA) Inc.
;      STREET:  3401 Hillview Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,620
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/224,917
;      FILING DATE:  08-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Peries, Rohan
;      REGISTRATION NUMBER:  35,752
;      REFERENCE/DOCKET NUMBER:  28060
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)-852-1698
;      TELEFAX:  (415)-496-3529
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2237 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  A3.01
;    IMMEDIATE SOURCE:
;      CLONE:  GMPS.6
US-08-463-620-1


  Query Match             48.0%;  Score 12;  DB 2;  Length 2237;
  Best Local Similarity   52.2%;  Pred. No. 65;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db   2157 GAAGATTCCTGGTATTTCTCGAA 2179


RESULT   6
US-08-224-917-1
; Sequence 1, Application US/08224917
; Patent No. 5965350
;  GENERAL INFORMATION:
;    APPLICANT:  Lou, Lillian Lien-Li
;    APPLICANT:  Barnett, Jimmy Wayne
;    TITLE OF INVENTION:  Cloning and Expression of Human GMP
;    TITLE OF INVENTION:  Synthetase, its use in Screening for Inhibitors of Human
;    TITLE OF INVENTION:  GMP Synthetase and Inhibitors of Human GMP Synthetase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Syntex (USA) Inc.
;      STREET:  3401 Hillview Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/224,917
;      FILING DATE:  08-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Peries, Rohan
;      REGISTRATION NUMBER:  35,752
;      REFERENCE/DOCKET NUMBER:  28060
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)-852-1698
;      TELEFAX:  (415)-496-3529
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2237 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  A3.01
;    IMMEDIATE SOURCE:
;      CLONE:  GMPS.6
US-08-224-917-1
```

Query Match 48.0%; Score 12; DB 4; Length 2237;
Best Local Similarity 52.2%; Pred. No. 65;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db   2157 GAAGATTCCTGGTATTTCTCGAA 2179
```

```
RESULT   7
US-08-914-853-1
; Sequence 1, Application US/08914853
; Patent No. 5998186
;  GENERAL INFORMATION:
;    APPLICANT:  Lou, Lillian Lien-Li
;    APPLICANT:  Barnett, Jimmy Wayne
;    TITLE OF INVENTION:  Cloning and Expression of Human GMP
;    TITLE OF INVENTION:  Synthetase, its use in Screening for Inhibitors of
;    TITLE OF INVENTION:  Human
;    TITLE OF INVENTION:  GMP Synthetase and Inhibitors of Human GMP Synthetase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Syntex (USA) Inc.
;      STREET:  3401 Hillview Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/914,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/461,489
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Peries, Rohan
;      REGISTRATION NUMBER:  35,752
;      REFERENCE/DOCKET NUMBER:  28060
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)-852-1698
;      TELEFAX:  (415)-496-3529
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2237 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  A3.01
;    IMMEDIATE SOURCE:
;      CLONE:  GMPS.6
US-08-914-853-1
```

Query Match 48.0%; Score 12; DB 4; Length 2237;
Best Local Similarity 52.2%; Pred. No. 65;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db   2157 GAAGATTCCTGGTATTTCTCGAA 2179
```

```
RESULT   8
PCT-US95-03934A-1
; Sequence 1, Application PC/TUS9503934A
;  GENERAL INFORMATION:
;    APPLICANT:  Syntex (USA) Inc.
;    TITLE OF INVENTION:  Cloning and Expression of Human GMP
;    TITLE OF INVENTION:  Synthetase, its use in Screening for Inhibitors of Human
;    TITLE OF INVENTION:  GMP Synthetase and Inhibitors of Human GMP Synthetase
;    NUMBER OF SEQUENCES:  11
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03934A
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2237 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  A3.01
;    IMMEDIATE SOURCE:
;      CLONE:  GMPS.6
PCT-US95-03934A-1
```

Query Match 48.0%; Score 12; DB 6; Length 2237;
Best Local Similarity 52.2%; Pred. No. 65;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db   2157 GAAGATTCCTGGTATTTCTCGAA 2179
```

```
RESULT   9
US-08-696-944-1/c
; Sequence 1, Application US/08696944
; Patent No. 5981831
;  GENERAL INFORMATION:
;    APPLICANT:  Sumant CHENGAPPA
;    APPLICANT:  Susan A. HELLYER
;    APPLICANT:  John S. REID
;    APPLICANT:  Jacqueline DE SILVA
;    TITLE OF INVENTION:  No. 5981831el Exo-(1-4)-Beta-D Galactanase
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,944
;      FILING DATE:  23-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB95/00372
;      FILING DATE:  23-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9403423.8
```

```
;      FILING DATE:  23-FEB-1994
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2628 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  130..2319
US-08-696-944-1


  Query Match             48.0%;  Score 12;  DB 4;  Length 2628;
  Best Local Similarity   52.2%;  Pred. No. 67;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||        |||  |||
Db     407 GAAGGTTCATGTCCATTCCAGAA 385


RESULT  10
US-08-433-126A-68/c
; Sequence 68, Application US/08433126A
; Patent No. 5688935
;  GENERAL INFORMATION:
;    APPLICANT:  STEPHENS, ANDREW
;    APPLICANT:  SCHNEIDER, DAN
;    APPLICANT:  GOLD, LARRY
;    TITLE OF INVENTION:  NUCLEIC ACID LIGANDS OF TISSUE
;    TITLE OF INVENTION:  TARGET
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Avenue, Suite 200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3 1/2 diskette, 1.44 MG
;      COMPUTER:  IBM pc compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,126A
;      FILING DATE:  03 MAY 1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/714,131
;      FILING DATE:  10-JUNE-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/536,428
;      FILING DATE:  11-JUNE-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/964,624
;      FILING DATE:  21-OCTOBER-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX31.2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  68:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  88 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  All C's are 2'-F cytosine
;    FEATURE:
;      OTHER INFORMATION:  All U's are 2'-F uracil
US-08-433-126A-68


  Query Match             44.0%;  Score 11;  DB 1;  Length 88;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+02;
  Matches   11;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnnga 23
           |||  |||  .     |||  ||
Db      50 GAATGTTCGGACTTCTTCAAGA 29


RESULT  11
US-08-433-124A-68/c
; Sequence 68, Application US/08433124A
; Patent No. 5750342
;  GENERAL INFORMATION:
;    APPLICANT:  STEPHENS, ANDREW
;    APPLICANT:  SCHNEIDER, DAN
;    APPLICANT:  GOLD, LARRY
;    TITLE OF INVENTION:  NUCLEIC ACID LIGANDS OF TISSUE
;    TITLE OF INVENTION:  TARGET
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Avenue, Suite 200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3 1/2 diskette, 1.44 MG
;      COMPUTER:  IBM pc compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,124A
;      FILING DATE:  03 MAY 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/714,131
;      FILING DATE:  10-JUNE-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/536,428
;      FILING DATE:  11-JUNE-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/964,624
;      FILING DATE:  21-OCTOBER-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX31.2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  68:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  88 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  All C's are 2'-F cytosine
;    FEATURE:
;      OTHER INFORMATION:  All U's are 2'-F uracil
US-08-433-124A-68
```

```
  Query Match             44.0%;  Score 11;  DB 2;  Length 88;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+02;
  Matches   11;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnnga 23
           |||  |||        |||  ||
Db      50 GAATGTTCGGACTTCTTCAAGA 29


RESULT  12
PCT-US96-06059-68/c
; Sequence 68, Application PC/TUS9606059
;  GENERAL INFORMATION:
;    APPLICANT:  STEPHENS, ANDREW
;    APPLICANT:  SCHNEIDER, DAN
;    APPLICANT:  GOLD, LARRY
;    TITLE OF INVENTION:  NUCLEIC ACID LIGANDS OF TISSUE
;    TITLE OF INVENTION:  TARGET
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Avenue, Suite 200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3 1/2 diskette, 1.44 MG
;      COMPUTER:  IBM pc compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06059
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/433,124
;      FILING DATE:  03-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/433,126
;      FILING DATE:  03-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/714,131
;      FILING DATE:  10-JUNE-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/536,428
;      FILING DATE:  11-JUNE-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/964,624
;      FILING DATE:  21-OCTOBER-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX31.2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  68:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  88 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  All C's are 2'-F cytosine
;    FEATURE:
;      OTHER INFORMATION:  All U's are 2'-F uracil
PCT-US96-06059-68


  Query Match             44.0%;  Score 11;  DB 6;  Length 88;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+02;
  Matches   11;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnnga 23
           |||  |||        |||  ||
Db      50 GAATGTTCGGACTTCTTCAAGA 29


RESULT  13
US-08-659-188-24/c
; Sequence 24, Application US/08659188
; Patent No. 6002069
;  GENERAL INFORMATION:
;    APPLICANT:  Yanofsky, Martin F.
;    TITLE OF INVENTION:  Seed Plants Exhibiting Inducible Early
;    TITLE OF INVENTION:  Reproductive Development and Methods of Making Same
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,188
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 1946
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  371 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..371
;      OTHER INFORMATION:  /note= "element = heat shock
;      OTHER INFORMATION:  inducible regulatory element (HSP81-1 promoter)."
US-08-659-188-24


  Query Match             44.0%;  Score 11;  DB 5;  Length 371;
  Best Local Similarity   50.0%;  Pred. No. 2e+02;
  Matches   11;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnnga 23
           |||  |||        |||  ||
Db      28 GAAAGTTCTTTTCGTTTCGAGA 7


RESULT  14
US-08-650-275-12/c
; Sequence 12, Application US/08650275
; Patent No. 5798249
;  GENERAL INFORMATION:
;    APPLICANT:  Braxton, Scott Michael
;    APPLICANT:  Murry, Lynn E.
;    TITLE OF INVENTION:  HUMAN PROTEIN DISULFIDE ISOMERASE
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
```

```
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  U.S.
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/650,275
;       FILING DATE:  Filed Herewith
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Luther, Barbara J.
;       REGISTRATION NUMBER:  33,954
;       REFERENCE/DOCKET NUMBER:  PF-0067 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-852-0195
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  394 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       LIBRARY:  LVENNOT01
;       CLONE:  350290
US-08-650-275-12


  Query Match             44.0%;  Score 11;  DB 2;  Length 394;
  Best Local Similarity   47.8%;  Pred. No. 2e+02;
  Matches   11;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||   ||
Db     316 GAATATTCCTAAACTTTCTGNAA 294


RESULT  15
US-09-181-318-12/c
; Sequence 12, Application US/09181318
; Patent No. 6001632
;  GENERAL INFORMATION:
;    APPLICANT:  Braxton, Scott Michael
;    APPLICANT:  Murry, Lynn E.
;    TITLE OF INVENTION:  HUMAN PROTEIN DISULFIDE ISOMERASE
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/181,318
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/650,275
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luther, Barbara J.
;      REGISTRATION NUMBER:  33,954
;      REFERENCE/DOCKET NUMBER:  PF-0067 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-852-0195
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  394 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      LIBRARY:  LVENNOT01
;      CLONE:  350290
US-09-181-318-12


  Query Match             44.0%;  Score 11;  DB 5;  Length 394;
  Best Local Similarity   47.8%;  Pred. No. 2e+02;
  Matches   11;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||   ||
Db     316 GAATATTCCTAAACTTTCTGNAA 294


Search completed: March  7, 2000, 00:19:55
Job time: 521 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2000, 21:20:27 ; Search time 97.27 Seconds
(without alignments)
64.304 Million cell updates/sec

Title: US-09-304-121-3
Perfect score: 25
Sequence: 1 ngaannttcnnnnnnnttcnngaan 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	48.0	85	1	N71047	Mutant SE7. Induci
c	2	12	48.0	85	1	N71047	Mutant SE7. Induci
	3	12	48.0	136	1	T22546	Human gene signatu
	4	12	48.0	152	1	V10679	Human breast cance
c	5	12	48.0	1070	1	Q05965	Sequence encoding
	6	12	48.0	1087	1	V68802	Human endogenous r
	7	12	48.0	1163	1	T75004	Human endogenous r
	8	12	48.0	1350	1	V22746	Babesia microti BM
	9	12	48.0	1410	1	V25088	H. pylori secreted
c	10	12	48.0	1785	1	Q12005	Human TR2-9 DNA bi
c	11	12	48.0	1839	1	T02337	Marek's disease ty
c	12	12	48.0	2029	1	Q12004	Human TR2-5 androg
c	13	12	48.0	2221	1	Q12006	Human TR2-11 DNA b
	14	12	48.0	2237	1	T00492	Human guanosine 5'
	15	12	48.0	2237	1	T00493	Human guanosine 5'
c	16	12	48.0	2458	1	Q12003	Human TR2-7 DNA bi
c	17	12	48.0	2628	1	T01014	Lupin exo-(1-4)bet
	18	12	48.0	3452	1	T58243	CH1-9a11-2, over a
	19	12	48.0	3452	1	V10689	Human breast cance
	20	12	48.0	5386	1	V10696	Human 3.5 kB DNA f
	21	12	48.0	6413	1	V31988	Human Down syndrom
	22	12	48.0	6604	1	V31981	Human Down syndrom
c	23	12	48.0	8084	1	X13109	Enterococcus faeca
c	24	12	48.0	14654	1	V52239	Streptococcus pneu
c	25	11	44.0	41	1	T70642	Fibrin clot bindin
	26	11	44.0	135	1	T23461	Human gene signatu
	27	11	44.0	153	1	T26747	Human gene signatu
c	28	11	44.0	214	1	T20153	Human gene signatu
	29	11	44.0	232	1	T24841	Human gene signatu
c	30	11	44.0	255	1	Q57451	Oxysterol binding
	31	11	44.0	275	1	V90413	EST clone DL634. N
c	32	11	44.0	294	1	Q39819	Expressed Sequence
c	33	11	44.0	294	1	Q59231	Human brain Expres
	34	11	44.0	302	1	V86842	EST clone BD151. N
c	35	11	44.0	328	1	V87865	EST clone ED205. N
c	36	11	44.0	329	1	N70070	Plasmid pHStPA. He
	37	11	44.0	347	1	V86470	EST clone AM889. N
	38	11	44.0	367	1	V15435	Human gene fragmen
c	39	11	44.0	371	1	V06028	Heat shock inducib
c	40	11	44.0	371	1	V02773	Heat shock inducib
c	41	11	44.0	371	1	T86663	Heat shock inducib
c	42	11	44.0	394	1	V09959	Partial cDNA Incyt
	43	11	44.0	461	1	V90037	EST clone DA443. N
c	44	11	44.0	500	1	V53304	DNA encoding a Sta
	45	11	44.0	544	1	V02706	Human Class I HLA-

ALIGNMENTS

RESULT 1
N71047
ID N71047 standard; DNA; 85 BP.
AC N71047;
DT 18-APR-1991 (first entry)
DE Mutant SE7.
KW Mutant SE7; synthetic linker; D50 DNA; promoter.
FH Key Location/Qualifiers
FT misc_feature 15. .18
FT /*tag= a
FT /note="consensus sequence"
FT misc_feature 23. .28
FT /*tag= b
FT /note="consensus sequence"
FT misc_feature 33. .38
FT /*tag= c
FT /note="consensus sequence"
FT misc_feature 43. .48
FT /*tag= d
FT /note="consensus sequence"
FT misc_feature 53. .58
FT /*tag= e
FT /note="consensus sequence"
FT misc_feature 63. .68
FT /*tag= f
FT /note="consensus sequence"
FT misc_feature 73. .75
FT /*tag= g
FT /note="consensus sequence"
PN WO8700861-A.
PD 12-FEB-1987.
PF 29-JUL-1986; E00451.
PR 31-JUL-1985; EP-810354.
PR 29-JUL-1986; EP-905251.
PR 29-JUL-1986; WO-E00451.
PR 26-MAR-1987; DK-001541.
PA (BATT) BATTELLE MEMORIAL INST.
PA (BROM/) BROMLEY P.
PI Bromley P, Dreano M, Voellmy R;
DR WPI; 87-050099/07.
PT Inducing expression of eukaryotic genes - using recombinant DNA
PT gene expression unit under control of heat-shock control element
PS Disclosure; page 20; 39pp; English.
CC The mutant was prepd. by preligating the synthetic linker
CC AGAAGCTTCT and ligating it to Xho I digested and blunt ended D50
CC DNA. It can be used in an expression system to give increased
CC levels of prodn. of a desired gene.
SQ Sequence 85 BP; 24 A; 19 C; 19 G; 23 T;

Query Match 48.0%; Score 12; DB 1; Length 85;
Best Local Similarity 52.2%; Pred. No. 93;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy       2 gaannttcnnnnnnnttcnngaa 24
           ||| |||        ||| |||
Db       8 GAAGCTTCTAGAAGCTTCTAGAA 30
```

```
RESULT   2
N71047/c
ID   N71047 standard; DNA; 85 BP.
AC   N71047;
DT   18-APR-1991  (first entry)
DE   Mutant SE7.
KW   Mutant SE7; synthetic linker; D50 DNA; promoter.
FH   Key             Location/Qualifiers
FT   misc_feature    15. .18
FT                   /*tag= a
FT                   /note="consensus sequence"
FT   misc_feature    23. .28
FT                   /*tag= b
FT                   /note="consensus sequence"
FT   misc_feature    33. .38
FT                   /*tag= c
FT                   /note="consensus sequence"
FT   misc_feature    43. .48
FT                   /*tag= d
FT                   /note="consensus sequence"
FT   misc_feature    53. .58
FT                   /*tag= e
FT                   /note="consensus sequence"
FT   misc_feature    63. .68
FT                   /*tag= f
FT                   /note="consensus sequence"
FT   misc_feature    73. .75
FT                   /*tag= g
FT                   /note="consensus sequence"
PN   WO8700861-A.
PD   12-FEB-1987.
PF   29-JUL-1986; E00451.
PR   31-JUL-1985; EP-810354.
PR   29-JUL-1986; EP-905251.
PR   29-JUL-1986; WO-E00451.
PR   26-MAR-1987; DK-001541.
PA   (BATT ) BATTELLE MEMORIAL INST.
PA   (BROM/) BROMLEY P.
PI   Bromley P, Dreano M, Voellmy R;
DR   WPI; 87-050099/07.
PT   Inducing expression of eukaryotic genes - using recombinant DNA
PT   gene expression unit under control of heat-shock control element
PS   Disclosure; page 20; 39pp; English.
CC   The mutant was prepd. by preligating the synthetic linker
CC   AGAAGCTTCT and ligating it to Xho I digested and blunt ended D50
CC   DNA.  It can be used in an expression system to give increased
CC   levels of prodn. of a desired gene.
SQ   Sequence   85 BP;    24 A;    19 C;    19 G;    23 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 85;
  Best Local Similarity   52.2%;  Pred. No. 93;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     75 GAAGCTTCTAGAAGCTTCTAGAA 53


RESULT   3
T22546
ID   T22546 standard; cDNA to mRNA; 136 BP.
AC   T22546;
DT   01-OCT-1996  (first entry)
DE   Human gene signature HUMGS04159.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1154; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   136 BP;    43 A;    22 C;    29 G;    42 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 136;
  Best Local Similarity   52.2%;  Pred. No. 1e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     43 GAAGATTCCTGGTATTTCTCGAA 65


RESULT   4
V10679
ID   V10679 standard; DNA; 152 BP.
AC   V10679;
DT   21-JUL-1998  (first entry)
DE   Human breast cancer gene CH1-9a11-2 DNA fragment.
KW   Breast cancer; CH1-9a11-2; malignant transformation; diagnostic;
KW   therapeutic; screening; ds.
OS   Homo sapiens.
PN   WO9738085-A2.
PD   16-OCT-1997.
PF   09-APR-1997; U05930.
PR   10-JUL-1996; US-678280.
PR   09-APR-1996; US-015167.
PR   05-JUN-1996; WO-U09286.
PR   06-JUN-1996; US-019202.
PA   (CALP-) CALIFORNIA PACIFIC MEDICAL CENT RES INST.
PI   Chen L, Smith H;
DR   WPI; 97-512705/47.
DR   P-PSDB; W40366.
PT   Breast cancer genes - used to develop products to design or screen
PT   diagnostic reagents or therapeutic compounds
PS   Claim 16; Fig 22; 118pp; English.
CC   This sequence encodes a fragment of a novel human breast cancer gene,
CC   CH1-9a11-2. This gene fragment can be used for identifying genes and
CC   gene products that are intimately related to malignant transformation or
CC   maintenance of the malignant properties of cancer cells. It can
CC   also be used to design or screen diagnostic reagents or therapeutic
CC   compounds. Kits are included within the scope of the invention.
SQ   Sequence   152 BP;    62 A;    32 C;    32 G;    26 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 152;
  Best Local Similarity   52.2%;  Pred. No. 1e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;
```

```
Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db     47 GAAACTTCAGAATACTTCAAGAA 69


RESULT   5
Q05965/c
ID   Q05965 standard; DNA; 1070 BP.
AC   Q05965;
DT   14-JAN-1991  (first entry)
DE   Sequence encoding rat interleukin-6 (IL-6).
KW   Immunostimulant; antitumour; antiinflamatory; cytokine; ds.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   cds             38. .670
FT                   /*tag=  a
PN   J02195885-A.
PD   02-AUG-1990.
PF   25-JAN-1989; 016806.
PR   25-JAN-1989; JP-016806.
PA   (SAKA ) OTSUKA PHARM KK.
DR   WPI; 90-278846/37.
DR   P-PSDB; R06847.
PT   Rat IL-6 gene - used in development of IL-6 for drugs e.g.
PT   immunostimulant, antitumour drug, antiinflammatory drug etc.
PS   Disclosure; Page 556; 16pp; Japanese.
CC   Sequence may be used to produce IL-6 useful in study and development
CC   of drugs eg. immunostimulants, antitumour drugs, cytokine production
CC   accelerators, antiinflamatory drugs and radiation damage inhibitors.
SQ   Sequence   1070 BP;    357 A;    195 C;    185 G;    333 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1070;
  Best Local Similarity  52.2%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db     47 GAAACTTCATAGCTGTTCCTGAA 25


RESULT   6
V68802
ID   V68802 standard; DNA; 1087 BP.
AC   V68802;
DT   22-JAN-1999  (first entry)
DE   Human endogenous retroviral DNA sequence #2.
KW   Human; breast cancer; breast tumour tissue; diagnosis; treatment;
KW   vaccine; epitope; endogenous; retroviral element; ss.
OS   Human endogenous retrovirus.
PN   WO9845328-A2.
PD   15-OCT-1998.
PF   09-APR-1998; U06939.
PR   11-DEC-1997; US-991789.
PR   09-APR-1997; US-838762.
PA   (CORI-) CORIXA CORP.
PI   Frudakis TN, Reed SG, Smith JM;
DR   WPI; 98-557473/47.
PT   New DNA sequences isolated from endogenous human retroviral element
PT   - and related vectors, transformed cells, proteins and antibodies,
PT   useful for diagnosis, treatment and prevention of breast cancer
PS   Claim 1; Page 36-37; 173pp; English.
CC   V68800 to V68998 represent nucleotide sequences which encode human
CC   breast tumour specific polypeptides. Detection or measurement of
CC   human breast tumour specific polypeptides and nucleotide sequences,
CC   or the corresponding RNA in a sample, is used for diagnosis and
CC   monitoring of breast cancer. Human breast tumour specific polypeptides
CC   and nucleotide sequences, and the vectors containing the DNAs, are also
CC   useful in vaccines for inhibiting development (for prevention or
CC   therapy) of breast cancer. The polypeptides may also be used to
CC   raise monoclonal antibodies, used as immunoassay reagents.
SQ   Sequence   1087 BP;    238 A;    266 C;    158 G;    363 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1087;
  Best Local Similarity  52.2%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db    513 GAATCTTCTTTAATGTTCTTGAA 535


RESULT   7
T75004
ID   T75004 standard; DNA; 1163 BP.
AC   T75004;
DT   06-OCT-1997  (first entry)
DE   Human endogenous retroviral sequence 11-29.
KW   Breast cancer; tumour; B18Ag1; prognosis; diagnosis; vaccine; ss.
OS   Human retrovirus.
PN   WO9725431-A1.
PD   17-JUL-1997.
PF   10-JAN-1997; U00398.
PR   10-JAN-1996; US-587329.
PA   (CORI-) CORIXA CORP.
PI   Frudakis TN,  Smith JM;
DR   WPI; 97-384982/35.
PT   Endogenous human tumour-associated retroviral element, B18Ag1 - used
PT   for the prognosis, diagnosis and monitoring of human cancers,
PT   especially breast cancer
PS   Claim 10; Page 28-29; 74pp; English.
CC   Human endogenous retroviral sequences 10, 11-29, 3, 6, 12, 13, 14
CC   and 11-22 (T75003-10) were obtd. by screening human genomic
CC   libraries using human breast tumour-associated retroviral element
CC   B18Ag1 (see also T75002) as probe.  These non-contiguous sequences
CC   lie in order 11-22, 14, B18Ag-1, 13, 12, 10, 3, 11-29, 6 in the
CC   retrovirus genome (see also T75001).  B18Ag1 and the other
CC   retroviral sequences can be used in genetic vaccines and for the
CC   prognosis, diagnosis and monitoring of human breast cancer.
SQ   Sequence   1163 BP;    257 A;    289 C;    176 G;    377 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1163;
  Best Local Similarity  52.2%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| |||       ||| |||
Db    589 GAATCTTCTTTAATGTTCTTGAA 611


RESULT   8
V22746
ID   V22746 standard; DNA; 1350 BP.
AC   V22746;
DT   28-SEP-1998  (first entry)
DE   Babesia microti BMNI-16 antigen sequence.
KW   antigen; detection; diagnosis; vaccine; tick-borne disease;
KW   differentiation; Lyme disease; ehrlichiosis; ss.
OS   Babesia microti.
FH   Key             Location/Qualifiers
FT   CDS             967. .1350
FT                   /*tag=  a
FT                   /product= antigen
PN   EP-834567-A2.
PD   08-APR-1998.
PF   01-OCT-1997; 117067.
PR   24-APR-1997; US-845258.
PR   01-OCT-1996; US-723142.
PA   (CORI-) CORIXA CORP.
PI   Houghton R, Lodes MJ, Reed SG, Sleath PR;
DR   WPI; 98-195465/18.
DR   P-PSDB; W56296.
PT   Polypeptides comprising Babesia microti antigens and their
```

```
PT   immunogenic fragments or epitopes - and related nucleic acid,
PT   vectors, transformed cells and antibodies, useful for diagnosis of
PT   infection and in protective vaccines
PS   Claim 8; Page 41-42; 113pp; English.
CC   The sequence is that encoding a polypeptide comprising at least
CC   one antigenic portion of a Babesia microti antigen. It can be used
CC   to diagnose B. microti infection by detecting specific antibodies
CC   in usual immunoassays. Infection can also be diagnosed using:
CC   (a) primers or probes derived from the coding sequence, in
CC   standard amplification or hybridisation tests, or (b) using
CC   antibodies to detect the corresponding antigen. It is also
CC   useful in vaccines to protect against infection, especially
CC   when formulated with an adjuvant. The new diagnostic methods
CC   allow rapid differentiation between B. microti infection and
CC   other tick-borne diseases (Lyme disease and ehrlichiosis) that
CC   have similar symptoms but require different treatments.
SQ   Sequence   1350 BP;    395 A;    319 C;    294 G;    342 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1350;
  Best Local Similarity   52.2%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    789 GAATCTTCAAATCGATTCTAGAA 811


RESULT   9
V25088
ID   V25088 standard; DNA; 1410 BP.
AC   V25088;
DT   07-JUL-1998  (first entry)
DE   H. pylori secreted protein ORF hp7p10290_25548812_f3_14.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   CDS             1. .1410
FT                   /*tag= a
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   P-PSDB; W55679.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claims 5,6,37; Pages 443-444; 1145pp; English.
CC   This sequence encodes a H. pylori secreted protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   1410 BP;    441 A;    257 C;    307 G;    405 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1410;
  Best Local Similarity   52.2%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    127 GAATATTCCAAACAATTCACGAA 149


RESULT  10
Q12005/c
ID   Q12005 standard; DNA; 1785 BP.
AC   Q12005;
DT   20-AUG-1991  (first entry)
DE   Human TR2-9 DNA binding protein coding sequence.
KW   TR2-type clone; DNA-binding protein; steroid hormone; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             127. .1530
FT                   /*tag= a
FT                   /product= TR2-9 receptor
PN   WO9107423-A.
PD   30-MAY-1991.
PF   19-OCT-1990; U06015.
PR   17-NOV-1989; US-438775.
PA   (ARCH-) ARCH DEV CORP.
PI   Liao S, Chang C;
DR   WPI; 91-178048/24.
DR   P-PSDB; R12227.
PT   Androgen receptor and TR2 DNA binding proteins - DNA sequences
PT   and antibodies for detection and quantification methods
PS   Claim 14; Fig 5; 79pp; English.
CC   TR2-9 receptor cDNA was isolated from a human prostate cDNA
CC   library. It is one of a number of TR2-type cDNA sequences which it
CC   is hoped will be used for isolation and structural analysis of
CC   other cellular receptors, their genes and ligands.
SQ   Sequence   1785 BP;    549 A;    391 C;    376 G;    469 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1785;
  Best Local Similarity   52.2%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    859 GAATATTCATGAACATTCCTGAA 837


RESULT  11
T02337/c
ID   T02337 standard; DNA; 1839 BP.
AC   T02337;
DT   23-MAY-1996  (first entry)
DE   Marek's disease type I virus S region inverted repeat DNA sequence.
KW   Marek's disease; type I virus; S region; inverted repeat; vaccine;
KW   recombinant production; viral vector; ds.
OS   Marek's disease type I virus.
PN   J07255488-A.
PD   09-OCT-1995.
PF   17-MAR-1994; 074315.
PR   17-MAR-1994; JP-074315.
PA   (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
DR   WPI; 95-378543/49.
PT   Recombinant Marek disease virus for expression of foreign gene -
PT   produced by recombination in viral inverted repeat region and useful
PT   as polyvalent vaccine or for admin. of active polypeptide(s) to fowl
```

```
PS   Example 2; Pages 8-9; 11pp; Japanese.
CC   A recombinant Marek's disease virus for the expression of a foreign
CC   gene, is prepd. by replacing the S region inverted repeat sequences
CC   T02337/38 with a foreign gene expression cassette. The cassette
CC   pref. comprises a gene fragment linked, downstream from a promoter,
CC   to a gene encoding an infection-preventative antigen derived from a
CC   pathogen other than Marek's disease type I virus. The recombinant
CC   virus can be used to provide polyvalent line vaccines, and for
CC   administering physiologically active substances, e.g. hormones, to
CC   fowl, esp. chickens.
SQ   Sequence   1839 BP;    513 A;    389 C;    469 G;    468 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 1839;
  Best Local Similarity  52.2%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db   1562 GAATTTTCGGACCAATTCACGAA 1540


RESULT  12
Q12004/c
ID   Q12004 standard; DNA; 2029 BP.
AC   Q12004;
DT   20-AUG-1991  (first entry)
DE   Human TR2-5 androgen receptor coding sequence.
KW   hAR; DNA-binding protein; steroid hormone; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             127. .1578
FT                   /*tag=  a
FT                   /product= TR2-5
FT                   /note= "calculated mol.wt.= 52,982"
PN   WO9107423-A.
PD   30-MAY-1991.
PF   19-OCT-1990; U06015.
PR   17-NOV-1989; US-438775.
PA   (ARCH-) ARCH DEV CORP.
PI   Liao S, Chang C;
DR   WPI; 91-178048/24.
DR   P-PSDB; R12224.
PT   Androgen receptor and TR2 DNA binding proteins - DNA sequences
PT   and antibodies for detection and quantification methods
PS   Claim 12; Fig 4; 79pp; English.
CC   This sequence was isolated by screening commercially available human
CC   testis and prostate lambda gt11 cDNA libraries. Initial screening
CC   was with probes designed for homology to nucleotide sequences in the
CC   DNA-binding domain of known steroid receptors. Positive clones were
CC   then screened with 24mer probes specific for the various steroid
CC   receptors to eliminate those which coded for known receptors. Any
CC   remaining clones were analysed by restriction mapping and were
CC   sequenced. Of 54 human testis clones identified as hAR coding
CC   sequences, 30 were classified as TR2-type and had sequences which
CC   overlapped to form a 2.1kb cDNA, including clone TR2-5.
SQ   Sequence   2029 BP;    623 A;    437 C;    412 G;    557 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 2029;
  Best Local Similarity  52.2%;  Pred. No. 1.6e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    859 GAATATTCATGAACATTCCTGAA 837


RESULT  13
Q12006/c
ID   Q12006 standard; DNA; 2221 BP.
AC   Q12006;
DT   20-AUG-1991  (first entry)
DE   Human TR2-11 DNA binding protein coding sequence.
KW   TR2-type clone; DNA-binding protein; steroid hormone; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             57. .1868
FT                   /*tag=  a
FT                   /product= TR2-11 receptor
FT   polya_signal    2180. .2185
FT                   /*tag=  b
PN   WO9107423-A.
PD   30-MAY-1991.
PF   19-OCT-1990; U06015.
PR   17-NOV-1989; US-438775.
PA   (ARCH-) ARCH DEV CORP.
PI   Liao S, Chang C;
DR   WPI; 91-178048/24.
DR   P-PSDB; R12228.
PT   Androgen receptor and TR2 DNA binding proteins - DNA sequences
PT   and antibodies for detection and quantification methods
PS   Claim 15; Fig 6; 79pp; English.
CC   TR2-11 receptor cDNA was isolated from a human prostate cDNA
CC   library. It is one of a number of TR2-type cDNA sequences which it
CC   is hoped will be used for isolation and structural analysis of
CC   other cellular receptors, their genes and ligands.
SQ   Sequence   2221 BP;    736 A;    446 C;    438 G;    601 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 2221;
  Best Local Similarity  52.2%;  Pred. No. 1.6e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    789 GAATATTCATGAACATTCCTGAA 767


RESULT  14
T00492
ID   T00492 standard; cDNA; 2237 BP.
AC   T00492;
DT   31-JAN-1996  (first entry)
DE   Human guanosine 5'-monophosphate synthetase from clone 6 (GMPS.6).
KW   Human guanosine 5'-monophosphate synthetase; GMPS.6; A3.01 cells; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             148. .2229
FT                   /*tag=  a
PN   WO9527789-A.
PD   19-OCT-1995.
PF   07-APR-1995; U03934.
PR   08-APR-1994; US-224917.
PA   (SYNT ) SYNTEX USA INC.
PI   Barnett JW,  Lou L;
DR   WPI; 95-366393/47.
DR   P-PSDB; R83123.
PT   New isolated human guanosine 5'-mono:phosphate synthetase - used to
PT   develop prods. for its study and for identifying inhibitors useful for
PT   e.g. anti-cancer or immunosuppressive therapy
PS   Claim 4; Page 30-31; 48pp; English.
CC   Naturally occurring human GMPS was purified from A3.01 cells and
CC   digested with trypsin. Nine tryptic peptides were resolved. Their
CC   sequences are given in R83124-R83132 and are indicated on R83122 FT.
CC   Based on the peptide sequences, degenerate oligos were synthesised
CC   in both the sense and antisense orientations and used in PCR. A
CC   fragment was generated with oligos 2S2 and 8A2 (see T00493 FT). 2S2
CC   & 8A2 corresp. to tryptic peptides 2 & 8. This PCR fragment (pcr.
CC   2S8A) (see T00493 FT) was used to screen an A3.01 cDNA library. The
CC   complete sequence of positive clone 6 (GMPS.6 T00492) was determined
CC   and is shown in Figure 1 (T00493). The derived AA sequence (R83123)
CC   of human GMP synthetase is shown in Figure 1 (R83122). The
CC   predicted mol. wt. of the enzyme - 76,725 - was in good agreement
CC   with the size indicated by polyacrylamide gel electrophoresis of
CC   the purified A3.01 human GMPS.
```

```
SQ   Sequence   2237 BP;   661 A;    469 C;   519 G;    588 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 2237;
  Best Local Similarity  52.2%;  Pred. No. 1.6e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db   2157 GAAGATTCCTGGTATTTCTCGAA 2179


RESULT  15
T00493
ID   T00493 standard; cDNA; 2237 BP.
AC   T00493;
DT   31-JAN-1996  (first entry)
DE   Human guanosine 5'-monophosphate synthetase from clone 6 (GMPS.6).
KW   Human guanosine 5'-monophosphate synthetase; GMPS.6; A3.01 cells; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             148. .2229
FT                   /*tag=  a
FT   misc_feature    928. .950
FT                   /*tag=  b
FT                   /label= primer 2S
FT   misc_feature    2077. .2099
FT                   /*tag=  c
FT                   /label= primer 8S
FT                   /note= "I think this should be 8A"
FT   misc_feature    928. .2099
FT                   /*tag=  d
FT                   /note= "corresp. to pcr.2S8A"
PN   WO9527789-A.
PD   19-OCT-1995.
PF   07-APR-1995; U03934.
PR   08-APR-1994; US-224917.
PA   (SYNT ) SYNTEX USA INC.
PI   Barnett JW,  Lou L;
DR   WPI; 95-366393/47.
DR   P-PSDB; R83122
PT   New isolated human guanosine 5'-mono:phosphate synthetase - used to
PT   develop prods. for its study and for identifying inhibitors useful for
PT   e.g. anti-cancer or immunosuppressive therapy
PS   Claim 4; Fig 1; 48pp; English.
CC   Naturally occurring human GMPS was purified from A3.01 cells and
CC   digested with trypsin. Nine tryptic peptides were resolved. Their
CC   sequences are given in R83124-R83132 and are indicated on R83122 FT.
CC   Based on the peptide sequences, degenerate oligos were synthesised
CC   in both the sense and antisense orientations and used in PCR. A
CC   fragment was generated with oligos 2S2 and 8A2 (see T00493 FT). 2S2
CC   & 8A2 corresp. to tryptic peptides 2 & 8. This PCR fragment (pcr.
CC   2S8A) (see T00493 FT) was used to screen an A3.01 cDNA library. The
CC   complete sequence of positive clone 6 (GMPS.6 T00492) was determined
CC   and is shown in Figure 1 (T00493). The derived AA sequence (R83123)
CC   of human GMP synthetase is shown in Figure 1 (R83122). The
CC   predicted mol. wt. of the enzyme - 76,725 - was in good agreement
CC   with the size indicated by polyacrylamide gel electrophoresis of
CC   the purified A3.01 human GMPS.
SQ   Sequence   2237 BP;   661 A;    469 C;   519 G;    588 T;


  Query Match            48.0%;  Score 12;  DB 1;  Length 2237;
  Best Local Similarity  52.2%;  Pred. No. 1.6e+02;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db   2157 GAAGATTCCTGGTATTTCTCGAA 2179


Search completed: March  6, 2000, 21:20:27

Job time: 3622 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2000, 23:49:41 ; Search time 1394.2 Seconds
(without alignments)
-54.447 Million cell updates/sec

Title: US-09-304-121-3
Perfect score: 25
Sequence: 1 ngaannttcnnnnnnnttcnngaan 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	48.0	40	34	S38922	S38922 {RcC9 repea
c	2	12	48.0	40	34	S38922	S38922 {RcC9 repea
	3	12	48.0	83	34	CEZK1017B	X61245 C.elegans r
c	4	12	48.0	83	34	CEZK1017B	X61245 C.elegans r
	5	12	48.0	85	5	A15908	A15908 hsp 70KDa g
c	6	12	48.0	85	5	A15908	A15908 hsp 70KDa g
	7	12	48.0	116	34	CELRCBC9B	L26115 Caenorhabdi
c	8	12	48.0	116	34	CELRCBC9B	L26115 Caenorhabdi
c	9	12	48.0	117	34	CELREPB	L26058 Caenorhabdi
	10	12	48.0	117	34	CEREPC9B	X12416 C. elegans
c	11	12	48.0	117	34	CEREPC9B	X12416 C. elegans
c	12	12	48.0	123	34	CEREPC9A	X12415 C. elegans
	13	12	48.0	142	34	CELRCBC9A	L26114 Caenorhabdi
c	14	12	48.0	142	34	CELRCBC9A	L26114 Caenorhabdi
	15	12	48.0	151	34	CEZK1043C	X61256 C.elegans r
c	16	12	48.0	151	34	CEZK1043C	X61256 C.elegans r
	17	12	48.0	152	5	AR016237	AR016237 Sequence
	18	12	48.0	163	34	CEZK1014A	X61241 C.elegans r
c	19	12	48.0	163	34	CEZK1014A	X61241 C.elegans r
	20	12	48.0	180	34	CEZK1043B	X61255 C.elegans r
c	21	12	48.0	180	34	CEZK1043B	X61255 C.elegans r
	22	12	48.0	242	34	CEREPC9I	X07681 Caenorhabdi
c	23	12	48.0	242	34	CEREPC9I	X07681 Caenorhabdi
c	24	12	48.0	250	13	G15386	G15386 human STS S
c	25	12	48.0	250	13	G15558	G15558 human STS S
c	26	12	48.0	250	13	G29233	G29233 human STS S
c	27	12	48.0	250	13	G33413	G33413 human STS S
c	28	12	48.0	274	35	AF113306	AF113306 Conus rat
	29	12	48.0	280	34	CEZK1085	X61261 C.elegans r
c	30	12	48.0	280	34	CEZK1085	X61261 C.elegans r
	31	12	48.0	285	34	CEZK1019A	X61247 C.elegans r
c	32	12	48.0	285	34	CEZK1019A	X61247 C.elegans r
	33	12	48.0	300	8	CNS019CY	AL111482 Botrytis
	34	12	48.0	303	8	S59778	S59778 18 kda heat
	35	12	48.0	303	8	S59779	S59779 18 kda heat
	36	12	48.0	303	13	HSA065ZB5	Z51828 H.sapiens (
c	37	12	48.0	309	13	HUMMATRY06	L22524 Human matri
c	38	12	48.0	322	13	HS210YG3	Z16959 H. sapiens
	39	12	48.0	332	34	CEZK1025A	X61249 C.elegans r
c	40	12	48.0	332	34	CEZK1025A	X61249 C.elegans r
	41	12	48.0	338	34	CEZK1027A	X61252 C.elegans r
c	42	12	48.0	338	34	CEZK1027A	X61252 C.elegans r
c	43	12	48.0	395	13	G50094	G50094 SHGC-82988
c	44	12	48.0	417	13	HSPE10G08	AL009432 H.sapiens
	45	12	48.0	419	34	CFU30220	U30220 Crithidia f

ALIGNMENTS

RESULT 1
S38922
LOCUS S38922 40 bp DNA INV 08-MAY-1993
DEFINITION {RcC9 repeat, clone ZK1019, repetitive element} {Caenorhabditis elegans, Genomic, 40 nt].
ACCESSION S38922
VERSION S38922.1 GI:250759

```
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular and genomic organization of clusters of repetitive DNA
            sequences in Caenorhabditis elegans
  JOURNAL   J. Mol. Biol. 226 (1), 159-168 (1992)
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 107962] from the original journal article.
            This sequence comes from Fig. 7.
            Map location: X.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Caenorhabditis elegans"
BASE COUNT       10 a      6 c      5 g     15 t      4 others
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 40;
  Best Local Similarity   65.2%;  Pred. No. 1e+03;
  Matches   15;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||| ||| |  |  |||  |||
Db     17 GAANATTCCNGANTTTTCTCGAA 39


RESULT   2
S38922/c
LOCUS       S38922         40 bp    DNA             INV       08-MAY-1993
DEFINITION  {RcC9 repeat, clone ZK1019, repetitive element} [Caenorhabditis
            elegans, Genomic, 40 nt].
ACCESSION   S38922
VERSION     S38922.1  GI:250759
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular and genomic organization of clusters of repetitive DNA
            sequences in Caenorhabditis elegans
  JOURNAL   J. Mol. Biol. 226 (1), 159-168 (1992)
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 107962] from the original journal article.
            This sequence comes from Fig. 7.
            Map location: X.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Caenorhabditis elegans"
BASE COUNT       10 a      6 c      5 g     15 t      4 others
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 40;
  Best Local Similarity   60.9%;  Pred. No. 1e+03;
  Matches   14;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          ||| ||||     | |||  |||
Db     24 GAATNTTCCAGAANTTTCTAGAA 2


RESULT   3
CEZK1017B
LOCUS       CEZK1017B      83 bp    DNA             INV       13-DEC-1994
DEFINITION  C.elegans repetitive DNA.
ACCESSION   X61245
VERSION     X61245.1  GI:6936
KEYWORDS    repetitive DNA.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 83)
  AUTHORS   La Volpe,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1991) A. La Volpe, CNR International Institute of
            Genetics and Biophysic, Via Marconi 10, 80125 Naples, ITALY
REFERENCE   2  (bases 1 to 83)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular evolotion of clusters of satellite-like DNA sequence in
            Caenorhabditis elegans
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 83)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular and genomic organization of clusters of repetitive DNA
            sequences in Caenorhabditis elegans
  JOURNAL   J. Mol. Biol. 226 (1), 159-168 (1992)
  MEDLINE   92318259
REFERENCE   4  (bases 1 to 83)
  AUTHORS   La Volpe,A.
  TITLE     A repetitive DNA family, conserved throughout the evolution of
            free-living nematodes
  JOURNAL   J. Mol. Evol. 39 (5), 473-477 (1994)
  MEDLINE   95106284
FEATURES             Location/Qualifiers
     source          1. .83
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone_lib="Lorist cosmids"
                     /clone="ZK1017"
     repeat_region   1. .83
                     /note="RcC9 repetitive sequence"
BASE COUNT       26 a     16 c     12 g     29 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 83;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||        |||  |||
Db     20 GAAATTTCCCGATTTTTCTAGAA 42


RESULT   4
CEZK1017B/c
LOCUS       CEZK1017B      83 bp    DNA             INV       13-DEC-1994
DEFINITION  C.elegans repetitive DNA.
ACCESSION   X61245
VERSION     X61245.1  GI:6936
KEYWORDS    repetitive DNA.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 83)
  AUTHORS   La Volpe,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1991) A. La Volpe, CNR International Institute of
            Genetics and Biophysic, Via Marconi 10, 80125 Naples, ITALY
REFERENCE   2  (bases 1 to 83)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
```

```
  TITLE     Molecular evolotion of clusters of satellite-like DNA sequence in
            Caenorhabditis elegans
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 83)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular and genomic organization of clusters of repetitive DNA
            sequences in Caenorhabditis elegans
  JOURNAL   J. Mol. Biol. 226 (1), 159-168 (1992)
  MEDLINE   92318259
REFERENCE   4  (bases 1 to 83)
  AUTHORS   La Volpe,A.
  TITLE     A repetitive DNA family, conserved throughout the evolution of
            free-living nematodes
  JOURNAL   J. Mol. Evol. 39 (5), 473-477 (1994)
  MEDLINE   95106284
FEATURES             Location/Qualifiers
     source          1. .83
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone_lib="Lorist cosmids"
                     /clone="ZK1017"
     repeat_region   1. .83
                     /note="RcC9 repetitive sequence"
BASE COUNT       26 a     16 c     12 g     29 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 83;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     57 GAAAATTCTGGAATGTTCCAGAA 45


RESULT   5
A15908
LOCUS       A15908         85 bp    DNA             PAT       17-FEB-1994
DEFINITION  hsp 70KDa gene including the transcriptional and translational
            site.
ACCESSION   A15908
VERSION     A15908.1  GI:488934
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 85)
  AUTHORS   .
  TITLE     AN IMPROVED HEAT-SHOCK CONTROL METHOD AND SYSTEM FOR THE PRODUCTION
            OF COMPETENT EUKARYOTIC GENE PRODUCTS
  JOURNAL   Patent: WO 8700861-A 9 12-FEB-1987;
FEATURES             Location/Qualifiers
     source          1. .85
                     /organism="Drosophila sp."
                     /db_xref="taxon:7242"
BASE COUNT       24 a     19 c     19 g     23 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 5;  Length 85;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db      8 GAAGCTTCTAGAAGCTTCTAGAA 30


RESULT   6
A15908/c
LOCUS       A15908         85 bp    DNA             PAT       17-FEB-1994
DEFINITION  hsp 70KDa gene including the transcriptional and translational
            site.
ACCESSION   A15908
VERSION     A15908.1  GI:488934
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 85)
  AUTHORS   .
  TITLE     AN IMPROVED HEAT-SHOCK CONTROL METHOD AND SYSTEM FOR THE PRODUCTION
            OF COMPETENT EUKARYOTIC GENE PRODUCTS
  JOURNAL   Patent: WO 8700861-A 9 12-FEB-1987;
FEATURES             Location/Qualifiers
     source          1. .85
                     /organism="Drosophila sp."
                     /db_xref="taxon:7242"
BASE COUNT       24 a     19 c     19 g     23 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 5;  Length 85;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     75 GAAGCTTCTAGAAGCTTCTAGAA 53


RESULT   7
CELRCBC9B
LOCUS       CELRCBC9B      116 bp    DNA             INV       18-NOV-1993
DEFINITION  Caenorhabditis briggsae (clone brid2) rcbC9 gene sequence.
ACCESSION   L26115
VERSION     L26115.1  GI:416347
KEYWORDS    .
SOURCE      Caenorhabditis briggsae DNA.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 116)
  AUTHORS   La Volpe,A.
  JOURNAL   Unpublished (1993)
FEATURES             Location/Qualifiers
     source          1. .116
                     /organism="Caenorhabditis briggsae"
                     /db_xref="taxon:6238"
BASE COUNT       52 a     24 c     14 g     26 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 116;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     25 GAAAATTCCAGAACATTCCAGAA 47


RESULT   8
CELRCBC9B/c
LOCUS       CELRCBC9B      116 bp    DNA             INV       18-NOV-1993
DEFINITION  Caenorhabditis briggsae (clone brid2) rcbC9 gene sequence.
ACCESSION   L26115
VERSION     L26115.1  GI:416347
```

```
KEYWORDS    .
SOURCE      Caenorhabditis briggsae DNA.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 116)
  AUTHORS   La Volpe,A.
  JOURNAL   Unpublished (1993)
FEATURES             Location/Qualifiers
     source          1. .116
                     /organism="Caenorhabditis briggsae"
                     /db_xref="taxon:6238"
BASE COUNT       52 a     24 c     14 g     26 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 116;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    112 GAATGTTCTGGATTTTTCTGGAA 90


RESULT   9
CELREPB/c
LOCUS       CELREPB       117 bp    DNA             INV       12-NOV-1993
DEFINITION  Caenorhabditis briggsae repeat region.
ACCESSION   L26058
VERSION     L26058.1  GI:415668
KEYWORDS    .
SOURCE      Caenorhabditis briggsae DNA.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   La Volpe,A.
  JOURNAL   Unpublished (1993)
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="Caenorhabditis briggsae"
                     /db_xref="taxon:6238"
     repeat_region   1. .117
                     /note="HSE-like consensus"
BASE COUNT       50 a     24 c     13 g     30 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 117;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     92 GAATGTTCTGGATCTTTCTGGAA 70


RESULT  10
CEREPC9B
LOCUS       CEREPC9B      117 bp    DNA             INV       31-AUG-1992
DEFINITION  C. elegans repetitive sequence C9 fragment b.
ACCESSION   X12416 X07683
VERSION     X12416.1  GI:6837
KEYWORDS    repetitive sequence.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   La Volpe,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1988) La Volpe A., CNR International Institute of
            Genetics and Biophysics, Via G. Marconi 10, 80125 Napoli, Italy
REFERENCE   2  (bases 1 to 117)
  AUTHORS   La Volpe,A., Ciaramella,M. and Bazzicalupo,P.
  TITLE     Structure, evolution and properties of a novel repetitive DNA
            family in Caenorhabditis elegans
  JOURNAL   Nucleic Acids Res. 16 (17), 8213-8231 (1988)
  MEDLINE   88335585
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="cosrC9"
                     /map="position 3, chromosome V"
     repeat_region   1. .117
                     /note="C9 fragment b"
BASE COUNT       41 a     20 c     24 g     32 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 117;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     67 GAACTTTCTAGAAAATTCGAGAA 89


RESULT  11
CEREPC9B/c
LOCUS       CEREPC9B      117 bp    DNA             INV       31-AUG-1992
DEFINITION  C. elegans repetitive sequence C9 fragment b.
ACCESSION   X12416 X07683
VERSION     X12416.1  GI:6837
KEYWORDS    repetitive sequence.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   La Volpe,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1988) La Volpe A., CNR International Institute of
            Genetics and Biophysics, Via G. Marconi 10, 80125 Napoli, Italy
REFERENCE   2  (bases 1 to 117)
  AUTHORS   La Volpe,A., Ciaramella,M. and Bazzicalupo,P.
  TITLE     Structure, evolution and properties of a novel repetitive DNA
            family in Caenorhabditis elegans
  JOURNAL   Nucleic Acids Res. 16 (17), 8213-8231 (1988)
  MEDLINE   88335585
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="cosrC9"
                     /map="position 3, chromosome V"
     repeat_region   1. .117
                     /note="C9 fragment b"
BASE COUNT       41 a     20 c     24 g     32 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 117;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    114 GAAAATTCTGGACCATTCCGGAA 92


RESULT  12
CEREPC9A/c
```

```
LOCUS       CEREPC9A      123 bp    DNA             INV       31-AUG-1992
DEFINITION  C. elegans repetitive sequence C9 fragment a.
ACCESSION   X12415 X07683
VERSION     X12415.1  GI:6836
KEYWORDS    repetitive sequence.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 123)
  AUTHORS   La Volpe,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1988) La Volpe A., CNR International Institute of
            Genetics and Biophysics, Via G. Marconi 10, 80125 Napoli, Italy
REFERENCE   2  (bases 1 to 123)
  AUTHORS   La Volpe,A., Ciaramella,M. and Bazzicalupo,P.
  TITLE     Structure, evolution and properties of a novel repetitive DNA
            family in Caenorhabditis elegans
  JOURNAL   Nucleic Acids Res. 16 (17), 8213-8231 (1988)
  MEDLINE   88335585
FEATURES             Location/Qualifiers
     source          1. .123
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="cosrC9"
                     /map="position 3, chromosome V"
     repeat_region   1. .123
                     /note="C9 fragment a"
BASE COUNT       45 a     21 c     22 g     35 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 123;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     82 GAAAATTCTGGACCATTCCGGAA 60


RESULT  13
CELRCBC9A
LOCUS       CELRCBC9A     142 bp    DNA             INV       18-NOV-1993
DEFINITION  Caenorhabditis briggsae (clone bridi1) rcbC9 gene sequence.
ACCESSION   L26114
VERSION     L26114.1  GI:416346
KEYWORDS    .
SOURCE      Caenorhabditis briggsae DNA.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 142)
  AUTHORS   La Volpe,A.
  JOURNAL   Unpublished (1993)
FEATURES             Location/Qualifiers
     source          1. .142
                     /organism="Caenorhabditis briggsae"
                     /db_xref="taxon:6238"
BASE COUNT       58 a     28 c     17 g     39 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 142;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db     74 GAACATTCCAGTTTTTTCCAGAA 96


RESULT  14
CELRCBC9A/c
LOCUS       CELRCBC9A     142 bp    DNA             INV       18-NOV-1993
DEFINITION  Caenorhabditis briggsae (clone bridi1) rcbC9 gene sequence.
ACCESSION   L26114
VERSION     L26114.1  GI:416346
KEYWORDS    .
SOURCE      Caenorhabditis briggsae DNA.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 142)
  AUTHORS   La Volpe,A.
  JOURNAL   Unpublished (1993)
FEATURES             Location/Qualifiers
     source          1. .142
                     /organism="Caenorhabditis briggsae"
                     /db_xref="taxon:6238"
BASE COUNT       58 a     28 c     17 g     39 t
ORIGIN


  Query Match             48.0%;  Score 12;  DB 34;  Length 142;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+03;
  Matches   12;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gaannttcnnnnnnnttcnngaa 24
          |||  |||       |||  |||
Db    121 GAATTTTCTAGAATTTTCTGGAA 99


RESULT  15
CEZK1043C
LOCUS       CEZK1043C     151 bp    DNA             INV       08-DEC-1992
DEFINITION  C.elegans repetitive DNA.
ACCESSION   X61256
VERSION     X61256.1  GI:6947
KEYWORDS    repetitive DNA.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 151)
  AUTHORS   La Volpe,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1991) A. La Volpe, CNR International Institute of
            Genetics and Biophysic, Via Marconi 10, 80125 Naples, ITALY
REFERENCE   2  (bases 1 to 151)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular evolotion of clusters of satellite-like DNA sequence in
            Caenorhabditis elegans
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 151)
  AUTHORS   Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La
            Volpe,A.
  TITLE     Molecular and genomic organization of clusters of repetitive DNA
            sequences in Caenorhabditis elegans
  JOURNAL   J. Mol. Biol. 226 (1), 159-168 (1992)
  MEDLINE   92318259
FEATURES             Location/Qualifiers
     source          1. .151
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone_lib="Lorist cosmids"
                     /clone="ZK1043"
     repeat_region   1. .151
                     /note="RcC9 repetitive sequence"
BASE COUNT       44 a     30 c     22 g     54 t      1 others
ORIGIN
```

Query Match 48.0%; Score 12; DB 34; Length 151;
Best Local Similarity 52.2%; Pred. No. 1.1e+03;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy       2 gaannttcnnnnnnnttcnngaa 24
           |||  |||       |||  |||
Db      25 GAACTTTCCCAATTTTTCTAGAA 47
```

Search completed: March 6, 2000, 23:49:43
Job time: 3591 sec

OM of: US-09-304-121-2 to: EST:* out_format : pfs

Date: Mar 7, 2000 1:19 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09304121/runat_06032000_151316_15703/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09304121
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-304-121-2
Query length: 529
Database: EST:*
Database sequences: 4538634
Database length: 1887831982
Search time (sec): 559.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est25:AI324484	+	1198.50	1729.13	1.5e-87	927	! AI324484 mo99b07.y1 Stratagene
gb_est30:AI628965	+	1185.00	1711.61	1.4e-86	763	! AI628965 ty79a02.x1 NCI_CGAP_Ki
gb_est38:AW054829	+	1129.00	1630.76	4.6e-82	730	! AW054829 ws60d05.x1 NCI_CGAP_Br
gb_est25:AI325062	-	1122.00	1617.65	2.4e-81	961	! AI325062 mo99b07.x1 Stratagene
gb_est26:AI393937	+	1108.00	1599.35	2.6e-80	795	! AI393937 tg11a08.x1 NCI_CGAP_CL
gb_est37:AW007349	+	1094.00	1580.15	3.0e-79	715	! AW007349 ws51d12.x1 NCI_CGAP_Br
gb_est35:AI809542	+	1090.00	1574.56	6.1e-79	701	! AI809542 wf30g01.x1 Soares_NFL_
gb_est30:AI634255	+	1089.00	1573.01	7.5e-79	707	! AI634255 tr84d01.x1 NCI_CGAP_Pa
gb_est30:AI651222	+	1085.00	1566.56	1.7e-78	751	! AI651222 wa98b09.x1 NCI_CGAP_GC
gb_est31:AI700961	+	1005.00	1452.24	4.0e-72	631	! AI700961 we09b01.x1 NCI_CGAP_Lu
gb_est28:AI521804	+	900.00	1300.32	1.2e-63	599	! AI521804 ti82f04.x1 NCI_CGAP_Ki
gb_est35:AI863994	+	876.00	1266.94	8.4e-62	522	! AI863994 wj54e02.x1 NCI_CGAP_Lu
gb_est17:AA619537	+	853.00	1232.41	7.0e-60	580	! AA619537 vo84h01.r1 Barstead mc
gb_est43:AW169960	+	846.00	1223.33	2.2e-59	524	! AW169960 xj35b05.x1 Soares_NFL_
gb_est37:AI934773	+	804.00	1162.90	5.2e-56	497	! AI934773 wp89c05.x1 NCI_CGAP_Br
gb_est31:AI692361	+	803.00	1160.43	7.2e-56	547	! AI692361 wd63e02.x1 NCI_CGAP_Lu
gb_est44:AW177014	+	801.50	1156.39	1.2e-55	651	! AW177014 CM3-CT0105-170899-004-
gb_est31:AI703424	+	785.00	1135.57	1.7e-54	485	! AI703424 we24a12.x1 NCI_CGAP_Lu
gb_est5:N34260	+	779.00	1126.18	5.8e-54	517	! N34260 yx79f03.r1 Soares melanc
gb_est9:AA118488	+	767.00	1108.22	5.8e-53	543	! AA118488 mo99b07.r1 Stratagene
gb_est19:AA808226	+	765.00	1107.33	6.5e-53	450	! AA808226 oc40f09.s1 NCI_CGAP_GC
gb_est5:N34263	+	753.00	1089.08	6.8e-52	486	! N34263 yx79f06.r1 Soares melanc
gb_est22:AI042312	+	751.00	1086.29	9.7e-52	481	! AI042312 oy13h07.x1 Soares_sene
gb_est12:AA291542	+	749.00	1082.20	1.6e-51	537	! AA291542 zt41e12.s1 Soares ovar
gb_est39:AW135472	+	737.00	1066.41	1.2e-50	461	! AW135472 UI-H-BI1-aca-f-02-0-UI
gb_est24:AI191359	+	732.50	1059.18	3.1e-50	492	! AI191359 qe32d05.s1 Soares_feta
gb_est39:AW137844	+	731.00	1057.65	3.8e-50	463	! AW137844 UI-H-BI1-adj-e-08-0-UI
gb_est23:AI152963	+	720.00	1038.80	4.3e-49	606	! AI152963 ud54d02.r1 Soares mous
gb_est19:AA807625	+	718.00	1039.36	4.0e-49	438	! AA807625 nv65h03.s1 NCI_CGAP_GC
gb_est18:AA721970	+	699.00	1009.42	1.9e-47	546	! AA721970 zh17b06.s1 Soares_pine
gb_est21:AA928989	+	691.00	998.40	7.6e-47	516	! AA928989 oo27e02.s1 NCI_CGAP_Lu
gb_est19:AA771801	+	678.00	979.30	8.8e-46	527	! AA771801 ai34e09.s1 Soares_para
gb_est7:W39659	-	677.50	980.60	7.5e-46	436	! W39659 zc19g03.r1 Soares_parath
gb_est18:AA713803	+	672.50	972.67	2.1e-45	464	! AA713803 nv70e05.s1 NCI_CGAP_GC
gb_est12:AA312294	+	667.00	965.62	5.1e-45	425	! AA312294 EST182991 Jurkat T-cel
gb_est15:AA457709	+	663.50	959.72	1.1e-44	459	! AA457709 zx87c07.s1 Soares ovar
gb_est23:AI122923	+	646.00	934.28	2.8e-43	460	! AI122923 ok31e05.s1 Soares_NSF_
gb_est19:AA743736	+	637.00	922.18	1.3e-42	420	! AA743736 ny91g12.s1 NCI_CGAP_GC
gb_est7:W62396	+	618.00	893.90	5.0e-41	448	! W62396 md98h04.r1 Soares mouse
gb_est38:AW016463	+	616.00	891.29	7.0e-41	436	! AW016463 UI-H-BI0p-aav-g-08-0-U
gb_est25:AI292013	+	580.50	838.93	5.8e-38	470	! AI292013 qm88c01.x1 NCI_CGAP_Lu
gb_est21:AA913874	+	550.00	797.61	1.2e-35	356	! AA913874 om21h05.s2 Soares_NFL_
gb_est3:R67111	+	518.00	750.11	5.1e-33	392	! R67111 yi31c07.r1 Soares pla
gb_est8:AA044583	+	492.00	707.09	1.3e-30	641	! AA044583 zk73g12.r1 Soares_p
gb_est3:R76298	+	478.00	693.83	7.0e-30	331	! R76298 yl23b10.r1 Soares bre

seq_name: gb_est25:AI324484

seq_documentation_block:
LOCUS AI324484 927 bp mRNA EST 23-DEC-1998
DEFINITION mo99b07.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone IMAGE:567829 5' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN); gb:X61753 M.musculus mRNA for heat shock transcription factor 1 (MOUSE);, mRNA sequence.
ACCESSION AI324484
VERSION AI324484.1 GI:4058913
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 927)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152840.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:342477
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 410.
FEATURES Location/Qualifiers
source 1. .927
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:567829"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 225 a 262 c 241 g 191 t 8 others
ORIGIN

alignment_scores:
Quality: 1198.50 Length: 309
Ratio: 4.220 Gaps: 6
Percent Similarity: 91.909 Percent Identity: 82.848

alignment_block:
US-09-304-121-2 x AI324484 ..

Align seg 1/1 to: AI324484 from: 1 to: 927

```
 28 AspProAspThrAspAlaLeuIleCysTrpSerProSerGlyAsnSerPh 44
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 12 GACCCGGACACAGACGCGCTCATCTGCTGGAGCCCGAGTGGGAACAGCTT 61

 44 eHisValPheAspGlnGlyGlnPheAlaLysGluValLeuProLysTyrP 61
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 62 CCACGTGTTTGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACT 111

 61 heLysHisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGly 77
    ||||||||||||||||||||||||||||||||||||||||||||||||||
112 TCAAGCACAACAACATGGCTAGCTTCGTGCGGCAGCTCAACATGTATGGC 161

 78 PheArgLysValValHisIleGluGlnGlyGlyLeuValLysProGluAr 94
    ||||||||||||||||||||||||||||||||||||||||||||||||||
162 TTCCGAAAAGTAGTCCACATTGAGCAGGGTGGCCTGGTCAAGCCTGAGAG 211

 94 gAspAspThrGluPheGlnHisProCysPheLeuArgGlyGlnGluGlnL 111
    ||||||||||||||||||||||||||||||||||||||||||||||||||
212 AGATGACACCGAGTTCCAGCATCCTTGTTTCTTGCGTGGACAGGAACAGC 261

111 euLeuGluAsnIleLysArgLysValThrSerValSerThrLeuLysSer 127
    ||||||||||||||||||||||||||||||||||||||||||||||||||
262 TCCTTGAGAACATCAAGAGGAAAGTGACCAGCGTGTCCACCCTGAAGAGT 311

128 GluAspIleLysIleArgGlnAspSerValThrLysLeuLeuThrAspVa 144
    ||||||||||||||||||||||||||||||||:::||||||||||||||||
312 GAGGACATAAAAATACGCCAGGACAGTGTCACCCGGCTGTTGACAGATGT 361

144 lGlnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeuLeuAlaM 161
    ||||||||||||||||||||||||||||||||||||||||||||||||||
362 GCAGCTGATGAAGGGGAAACAGGAGTGTATGGACTCCAAGCTNCTGGCCA 411

161 etLysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGln 177
    ||||||||||||||||||||||||||||||:::   ||||||   ||||||
412 TGAAGCACGAGAACGAGGCCCTGTGGCGGAGGGT.GCCAGCCCTCGGCAG 460

178 LysHisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIl 194
    ||||||||||||||||||||||||||||||||||||||||||||||||||
461 AAGCATGCCCAGCAGCAAAAAGTTGTCAACAAGCTCATTCAGTTCCTGAT 510

194 eSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleProL 211
    ||||||||||||||||||||||||||||   ||||||||||||||||||
511 CTCACTGGTGCAGTCGAACCGGATCCTGNGGGTGAAGAGAAAGATCCCTC 560

211 euMetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSerArg 227
    |||||   :::||||||:::||||||||||||:::||||||||||:::|||
561 TGATGTNGAGTGACAGCAACTCAGCACACTCTGTGCCCAAGTATGGTCGA 610

228 GlnPheSerLeuGluHisValHisGlySer.GlyProTyrSerAlaProS 244
    |||:::   |||||||||||||||||||| ||||||||||||||||||||
611 CAGTACTTCCTGGAGCATGTNCATGGGTCCTGGCCCATACTCAGCTCCAT 660

244 erProAlaTyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSer 260
    ||   ||||||||||||||||||||||      |||||||||:::|||
661 CTTCAGCCTACAGCAGCTCTAGCCTTTACTTCTCTGATGCTGTCACCAGC 710

261 SerGlyProIleIleSerAspIleThrGluLeuAlaProAlaSer.ProM 277
    |||   |||||||||:::|||||||||||||||||||||   ::: |||:
711 TCTTGACCCATAATCTNCGATATCACTGAGCTGGCTCCA...ACAGCCCT 757

277 etAlaSerProGlyGlySerIleAspGluArgProLeuSerSerSerPro 293
    :::::   |||:::   |||||||||||||:::|||||||:::|||||||
758 TTGGCCTCCCTAGCAGGAGCATAGATGAGAAGCCTCTGTNCAGCAGCACT 807

294 LeuValArgValLysGluGluProProSerProProGlnSerProArgVa 310
    |||   ||||||:::::::::|||||:::      |||:::      ::
808 CTGTCCCGTGTCAGCAAGAGCCCCCCAGNCAACCTACAGGCCTTCGGTAC 857

310 lGluGluAlaSerProGlyArgProSerSerValAspThrLeuLeu.Ser 326
    :      :::::::|||   |||||||||:::::::   |||   ||| :::
858 T...GGAGGCAAGCCCTGCCGGCCATCCTNCATGATAACC...CTTTGGC 901
```

```
    327 ProThrAlaLeuIleAspSer 333
        ||||||:::::::|||||||||
    902 CCAACTTGCTTTATTGATTCA 922

seq_name: gb_est30:AI628965

seq_documentation_block:
LOCUS       AI628965      763 bp    mRNA            EST       23-APR-1999
DEFINITION  ty79a02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2285258 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI628965
VERSION     AI628965.1  GI:4665765
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 763)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 16, 1998 this sequence version replaced gi:2961738.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .763
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2285258"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      175 a    236 c    219 g    123 t     10 others
ORIGIN


alignment_scores:
            Quality: 1185.00             Length:    252
              Ratio:  4.938               Gaps:      1
 Percent Similarity:  95.238   Percent Identity: 92.063

alignment_block:
 US-09-304-121-2 x AI628965   ..

 Align seg 1/1  to: AI628965  from: 1  to: 763

       2 AspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAlaPh 18
         ||||||||||||||||||||||||||||||||   ||||||||||||||
       2 GATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCATCAACGTCCCGGCCTT 51
```

```
 18 eLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaLeuI 35
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 52 CCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGCTCA 101

 35 leCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGlyGln 51
    ||||||||||||||||||||||||||||||||||||||||||||||||||
102 TCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGCCAG 151

 52 PheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAlaSe 68
    ||||||||||||||||||||||||||||||||||||||||||||||||||
152 TTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAG 201

 68 rPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisIleG 85
    ||||||||||||||||||||||||||||||||||||||||||||||||||
202 CTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCG 251

 85 luGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGlnHis 101
    ||||||||||||||||||||||||||||||||||||||||||||||||||
252 AGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCAC 301

102 ProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLy 118
    ||||||||||||||||||||||||||||||||||||||||||||||||||
302 CCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAGGAA 351

118 sValThrSerValSerThrLeuLysSerGluAspIleLysIleArgGlnA 135
    ||||||||||||||||||||||||||||||||||||||||||||||||||
352 AGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCCAGG 401

135 spSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLysGln 151
    ||||||||||||||||||||||||||||||||||||||||||||||||||
402 ACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAGCAG 451

152 GluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLe 168
    ||||||||||||||||||||||||||||||||||||||||||||||||||
452 GAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCT 501

168 uTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysV 185
    ||||||||||||||||||||||||||||||||||||||||||||||||||
502 GTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAG 551

185 alValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsnArg 201
    ||||||||||||||||||||||||||||||||||||||||||||||||||
552 TCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGA 601

202 IleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySe 218
              ::::::||||||||||||||||||||||||||||||||||
602 TNCCCTGNGNGTGAAGAGAAGATCCCCCTGATGCTGAACGACAGTGGCTC 651

218 rAlaHisSerMetProLysTyrSerArgGln.PheSerLeuGluHisVal 234
    ||||||||||||||||:::|||:::   :::    ||||||      |||
652 AGCACATTCCATGCCCCAGTATAAGNCCCGCAGTCTCCCTGGNAGACGTN 701

235 HisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSerSe 251
    |||   ||||||||||||||:::|||   ||||||||||::::::||
702 CACCGGTCGGGCCCCTACTCGGNCCCCCTCCCAGCCTACAGCAGNNTCAG 751

251 rLeu 252
    ||||
752 CCTN 755

seq_name: gb_est38:AW054829

seq_documentation_block:
LOCUS       AW054829      730 bp    mRNA            EST       23-SEP-1999
DEFINITION  ws60d05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2501577 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AW054829
VERSION     AW054829.1  GI:5920532
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187940.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2501577"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      167 a    215 c    223 g    125 t
ORIGIN


alignment_scores:
            Quality: 1129.00            Length:    223
              Ratio:  5.109               Gaps:      0
 Percent Similarity:  99.103   Percent Identity: 98.206

alignment_block:
 US-09-304-121-2 x AW054829   ..

 Align seg 1/1  to: AW054829  from: 1  to: 730

   1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  38 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 87

  17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  88 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 137

  34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 138 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 187

  51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 188 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 237

  67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 238 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 287
```

```
     84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    288 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 337

    101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    338 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 387

    117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    388 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 437

    134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    438 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 487

    151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    488 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 537

    167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    538 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 587

    184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
        ||||||||||||||||||||||:::||||||||||||||||||||||||
    588 AAGTCGTCAACAAGCTCATTCAGTTGCTGATCTCACTGGTGCAGTCAAAC 637

    201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
        ||||||||||||||||||||||||||   |||||||||   ||||||||
    638 CGGATCCTGGGGGTGAAGAGAAAGATCCTGCTGATGCTGTACGACAGTGG 687

    217 ySerAlaHisSerMetPro 223
        ||||||||||:::|||||||
    688 CTCAGCACATGCCATGCCG 706
```

seq_name: gb_est25:AI325062

seq_documentation_block:

```
LOCUS       AI325062      961 bp    mRNA            EST       23-DEC-1998
DEFINITION  mo99b07.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:567829 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1
            (HUMAN); gb:X61753 M.musculus mRNA for heat shock transcription
            factor 1 (MOUSE);, mRNA sequence.
ACCESSION   AI325062
VERSION     AI325062.1  GI:4059491
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 961)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151652.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:342477
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 427.
FEATURES              Location/Qualifiers
     source           1. .961
                      /organism="Mus musculus"
                      /strain="NIH/Swiss"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:567829"
                      /clone_lib="Stratagene mouse heart (#937316)"
                      /sex="pooled"
                      /tissue_type="heart"
                      /dev_stage="13 day embryos"
                      /lab_host="SOLR (kanamycin resistant)"
                      /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                      Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      190 a    227 c    314 g    221 t      9 others
ORIGIN
```

alignment_scores:

```
            Quality: 1122.00             Length:    302
              Ratio:  4.140                Gaps:      3
 Percent Similarity:  89.735   Percent Identity: 80.464
```

alignment_block:

US-09-304-121-2 x AI325062/rev ..

Align seg 1/1 to reverse of: AI325062 from: 1 to: 961

```
    131 LysIleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMe 147
        ||||||||||||||||||   |||   ||||||   |||||||||   ||
    958 AAAATACGCCAGGACAGTTTCACCGGGCTGTTGCAAGATGTGCAGTT.AT 910

    147 tLysGlyLysGlnGluCysMetAspSerLysLeuLeuAla.MetLysHis 163
        |||||||:::|||:::|||||||||   ||||||   ::: |||||||||
    909 GAAGGGAAACCAGAAGTGTATGGACTTCAAGCTTCNTGGACATGAAGCAC 860

    164 GluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAl 180
        ||||||:::::|||||||||:::|||||||||||||||||||||||||||
    859 GAGAACAAGGGCCTTTGGCGGAAGGTGGCAAGCCTCCGGCAGAAGCATGC 810

    180 aGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleSerLeuV 197
        |:::||||||||||||||||||||||||||||||||||||||||||||||
    809 CAGCCAGCAAAAAGTTGTCAACAAGCTCATTCAGTTCCTGATCTCACTGG 760

    197 alGlnSerAsnArgIleLeuGlyValLysArgLysIleProLeuMetLeu 213
        ||||||||||||||||||||||||||||||||||||   |||||||||
    759 TGCAGTCGAACCGGATTCTGGGGGTGAAGAGAAAGATCCTTCTGATGTTG 710

    214 AsnAspSerGlySerAlaHisSerMetProLysTyrSerArgGlnPheSe 230
        :::||||||:::|||||||||||||:::||||||||||:::||||||:::||
    709 AGTGACAGCAACTCAGCNCACTCTGTGCCCAAGTATGGTCGACAGTACTC 660

    230 rLeuGluHisValHisGlySerGlyProTyrSerAlaProSerProAlaT 247
        |      ||||||||||||:::|||||||||||||||||||||||||||||
    659 CTGNNAGCATGTCCATGGTCNTGGCCCATACTCAGCTCCATCTCCAGCNT 610

    247 yrSerSerSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyPro 263
        ||||||||||||||:::|||:::   |||||||||:::||||||||||||
    609 ACAGCAGCTCTAGCATTTACTCCTCTGATGCTGTCACCAGCTCTGGACCC 560

    264 IleIleSerAspIleThrGluLeuAlaProAlaSerProMetAlaSerPr 280
        ||||||||||||||||||||||||||   :::||||||   ||||||||
    559 ATAATCTCCGATATCACTGAGCTGGCTCNCACCAGCCCTNTGGCCTCNCC 510

    280 oGlyGlySerIleAspGluArgProLeuSerSerSerProLeuValArgV 297
        ||||   ||||||||||||||||||||||||||||   |||   ||||
    509 AGGCAGGAGCATAGATGAGAGGCCTCTGTCCAGCAGCACTCTGGT.CGTG 461

    297 alLysGluGluProProSerProProGlnSerProArgValGluGluAla 313
```

```
      |||||:::||||||   |||||||||:::||||||||||||   ||||||
  460 TCAAGCAAGAGCCCACCAGCCCACCTCACAGCCCTCGGGTACTGGAGGCG 411

  314 SerProGlyArgProSerSerValAspThrLeuLeuSerProThrAlaLe 330
      ||||||||||||||||||||:::||||||   ||||||||||||||::
  410 AGCCCTGGGCGCCCATCCTCCATGGATACCCCTTTGTCCCCAACTGCCTT 361

  330 uIleAspSerIleLeuArgGluSerGluProAlaPro...AlaSerValT 346
      :|||||||||||||||||||||||||||||||:::|||   ||||||   |
  360 CATTGACTCCATCCTTCGAGAGAGCGAGCCTACCCCTGCTGCCTCAAACA 311

  346 hrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPro 362
      |||||      |||:::   |||                 ::::::   |||
  310 CAGCCCCTATGGACACAACCGGA.............GCCCAAGCCCCC 276

  363 SerProProProThrSerThrProGluLysCysLeuSerValAlaCysLe 379
      :::   |||      ||||||||||||||||||||||||||||||||||
  275 GCACTCCCGACCCCCTCCACCCCTGAGAAGTGCCTCAGCGTAGCCTGCCT 226

  379 uAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuA 396
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  225 AGACAAGAACGAGCTAAGTGATCACCTGGATGCCATGGACTCCAACCTGG 176

  396 spAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThrSer 412
      ||||||||||||||||||:::||||||||||||||||||||||||||||
  175 ACAACCTGCAGACCATGCTGACAAGCCACGGCTTCAGTGTGGACACCAGT 126

  413 AlaLeuLeuAspLeuPheSerProSerValThrValProAspMetSerLe 429
      ||||||||||||||||||||||||||||||:::|||||||||||||||||
  125 GCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCATGCCCGACATGAGCCT 76

  429 uPro 430
      ||||
   75 GCCT 72

seq_name: gb_est26:AI393937

seq_documentation_block:
LOCUS       AI393937      795 bp    mRNA            EST       30-MAR-1999
DEFINITION  tg11a08.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2108438 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI393937
VERSION     AI393937.1  GI:4223484
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 795)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 5, 1998 this sequence version replaced gi:2747316.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 878   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .795
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2108438"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      176 a    249 c    234 g    136 t
ORIGIN


alignment_scores:
            Quality: 1108.00             Length:    255
              Ratio:  4.598                Gaps:      2
 Percent Similarity:  94.510   Percent Identity: 90.980

alignment_block:
 US-09-304-121-2 x AI393937   ..

 Align seg 1/1  to: AI393937  from: 1  to: 795

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      38 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 87

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 137

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     138 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 187

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 237

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     238 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 287

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     288 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 337

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     338 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 387

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     388 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 437

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     438 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 487

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 537

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||   :::||||||   |||:::||||||||||||  :
     538 TCTGTGGCGGGAGGGTGGCAGCCTTCCGCAGGAGCATGCCCAGCAACCAG 587

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
```

```
      ::   |||:::|||||||||||||||||||||||||||||||||||||
    588 AAGGCGTGCACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAAC 637

    201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
        ||||||||||||||||||||   ||||||||||||||||||||||||
    638 CCGATCCTGGGGGTGAAGAGAAAGAT.CCCCTGATGCTGAACGACAGTGG 686

    217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
        |   ||||||||||||||   ||||||||||   ||||||||||||
    687 CTA.GCACATTCCATGCCCATGTATAGCCGGCAGTC.TCCTTGGAGCACG 734

    234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
        ||||||||||:::|||   :::   |||   |||:::      ||||||
    735 TCCACGGATCGGCCCCT...ACTCGTCCACTGCCCAGCTAACGCAGCTCC 781

    251 SerLeuTyrAlaPro 255
        |||   ::::::|||
    782 AGC...TTTAGCCCC 793

seq_name: gb_est37:AW007349

seq_documentation_block:
LOCUS       AW007349      715 bp    mRNA            EST       10-SEP-1999
DEFINITION  ws51d12.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2500727 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AW007349
VERSION     AW007349.1  GI:5856127
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188853.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 489.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2500727"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      168 a    213 c    212 g    121 t      1 others

ORIGIN


alignment_scores:
            Quality: 1094.00              Length:    226
              Ratio:  4.973                 Gaps:      1
 Percent Similarity:  97.345   Percent Identity: 95.133

alignment_block:
 US-09-304-121-2 x AW007349   ..

 Align seg 1/1  to: AW007349  from: 1  to: 715

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      38 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 87

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 137

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     138 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 187

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 237

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     238 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 287

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     288 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 337

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     338 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 387

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     388 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 437

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     438 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 487

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 537

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     538 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 587

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSer.As 200
         ||||||||||||||||   |||   ||||||||||||||||||| ||
     588 AAGTCGTCAACAAGCTCAATCAGNTTCTGATCTCACTGGTGCAGTCAAAA 637

     200 nArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerG 217
         ||||:::||||||||||||:::||||||||||||||||||:::   ||||
     638 CCGATTCCTGGGGGTGAAGAAGAAGATCCCCCTGATGCTGAAACACAGTG 687

     217 lySerAlaHisSerMetProLysTyr 225
         ||||||||   |||:::   :::|||
     688 GCTCAGCACCTTCCATTGGCCAGTAT 713

seq_name: gb_est35:AI809542

seq_documentation_block:
LOCUS       AI809542      701 bp    mRNA            EST       07-JUL-1999
```

```
DEFINITION  wf30g01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2357136 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1
            (HUMAN);, mRNA sequence.
ACCESSION   AI809542
VERSION     AI809542.1  GI:5396108
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247218.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2357136"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      162 a    209 c    211 g    118 t      1 others
ORIGIN


alignment_scores:
            Quality: 1090.00            Length:    222
              Ratio:  5.023              Gaps:      1
 Percent Similarity:  97.748   Percent Identity: 96.396

alignment_block:
 US-09-304-121-2 x AI809542   ..

 Align seg 1/1  to: AI809542  from: 1  to: 701

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||   ||||||||||||||
      38 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGGCCAGCAACGTCCCGGC 87

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 137

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     138 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 187

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 237

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     238 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 287

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     288 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 337

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     338 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 387

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     388 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 437

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     438 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 487

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 537

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     538 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 587

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSer.As 200
         ||||||||||||||||||::::   |||||||||||||||||||| ||
     588 AAGTCGTCAACAAGCTCATTTCAGTTCTGATCTCACTGGTGCAGTCAAAA 637

     200 nArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerG 217
         |||||||   |||   ::::::||||||||||||||||||||||||||||
     638 CCGGATCNCTGGGGGTGAAGAGAAGATCCCCCTGATGCTGAACGACAGTG 687

     217 lySerAlaHisSer 221
         |||||||||||||
     688 GCTCAGCACATTCC 701

seq_name: gb_est30:AI634255

seq_documentation_block:
LOCUS       AI634255      707 bp    mRNA            EST       26-APR-1999
DEFINITION  tr84d01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224993 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI634255
VERSION     AI634255.1  GI:4685585
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138329.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Homo sapiens"
```

```
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2224993"
                    /clone_lib="NCI_CGAP_Pan1"
                    /tissue_type="adenocarcinoma"
                    /lab_host="DH10B"
                    /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                    Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                    Average insert size 1.72 kb. Life Technologies catalog #:
                    11548-013"
BASE COUNT      169 a    207 c    211 g    116 t      4 others
ORIGIN


alignment_scores:
            Quality: 1089.00             Length:    235
              Ratio:  4.840               Gaps:      1
 Percent Similarity:  95.745   Percent Identity: 91.915

alignment_block:
 US-09-304-121-2 x AI634255   ..

 Align seg 1/1  to: AI634255  from: 1  to: 707

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       2 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 51

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      52 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 101

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 151

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     152 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 201

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         |||||||||||||||||||||||||||||||||   ||||||||||||||
     202 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTTCCGAAAGTGGTCCACA 251

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 301

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 351

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     352 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 401

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     402 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 451

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||:::||||||||||||||||||||||||||||||||||
     452 CAGGAGTGCATGGACTNCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 501

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||   :::|||||||||||||||||||||||||:::||||
     502 TCTGTGGCGGGAGGGTGGCAGCCTTCGGCAGAAGCATGCCCAGCACCAGA 551

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
         ||||||||||||||||||||   ||||||||||||||||||||||||:::
     552 AAGTCGTCAACAAGCTCATTCAGGTCCTGATCTCACTGGTGCAGTCAAAA 601

     201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
                 :::::::::||||||||||||||||||||||||||||||||||
     602 CCGGATCCCTGGGCTGAAGAGAAGATCCCCCTGATGCTGAACGACAGTGG 651

     217 ySerAlaHisSerMetProLysTyrSer..ArgGlnPheSerLeuGluHi 233
         ||||||||||||||||||||:::|||:::  |||      ||||||||||
     652 CTCAGCACATTCCATGCCCCAGTATANAGACGCAGNTCTTCCCTGGAGCA 701

     233 sVal 234
         ||||
     702 CGTN 705

seq_name: gb_est30:AI651222

seq_documentation_block:
LOCUS       AI651222      751 bp    mRNA            EST       04-MAY-1999
DEFINITION  wa98b09.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2304185 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI651222
VERSION     AI651222.1  GI:4735201
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 20, 1998 this sequence version replaced gi:2979980.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 483.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2304185"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      170 a    228 c    222 g    125 t      6 others
ORIGIN


alignment_scores:
            Quality: 1085.00             Length:    237
              Ratio:  4.801               Gaps:      0
 Percent Similarity:  95.359   Percent Identity: 93.249

alignment_block:
 US-09-304-121-2 x AI651222   ..
```

```
Align seq 1/1  to: AI651222  from: 1  to: 751

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      38 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 87

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      88 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 137

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     138 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 187

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 237

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     238 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 287

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     288 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 337

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     338 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 387

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     388 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 437

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     438 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 487

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 537

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     538 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 587

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
           ||||||||||||||||||||   ||||||||||||||:::::::|||
     588 NAGTCGTCAACAAGCTCATTCAGTNTCTGATCTCACTGGTGCGTCAAAAC 637

     201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
         |||      |||   |||   ||||||||||||||||||:::   |||||
     638 CGATNNCCTGGGGGGAAGAG.AAGATCCCCCTGATGCTGAAACACAGTGG 686

     217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
         ||||||||||||||||   :::|||:::      |||||||||      |
     687 CTCAGCACATTCCATGGC.CAGTATAAGGCCGCATTCTCCCTGGNAGACG 735

     234 alHisGlySer 237
         ||||||||||
     736 TNCACGGCTCG 746

seq_name: gb_est31:AI700961

seq_documentation_block:
LOCUS       AI700961      631 bp    mRNA            EST       03-JUN-1999
DEFINITION  we09b01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2340553 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI700961
VERSION     AI700961.1  GI:4988861
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137377.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2340553"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      147 a    184 c    191 g    108 t      1 others
ORIGIN


alignment_scores:
            Quality: 1005.00             Length:    197
              Ratio:  5.154                Gaps:      0
 Percent Similarity:  98.985   Percent Identity: 98.985

alignment_block:
 US-09-304-121-2 x AI700961   ..

 Align seq 1/1  to: AI700961  from: 1  to: 631

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      20 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 69

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      70 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 119

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     120 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 169

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     170 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 219

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     220 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 269

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     270 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 319

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     320 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 369

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     370 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 419

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     420 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 469

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     470 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 519

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||      ||||||||||||||||||||||||||||||||||||||||
     520 TCTNGTGGCGGAGGTGGCCAGCCTTCGGCAGAAGCATGCTCAGCAACAGA 569

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuVal 197
         |||||||||||||||||||||||||||||||||||||||
     570 AAGTCGTCAACAAGCTCATTCAGTTTCTGATCTCACTGGTG 610

seq_name: gb_est28:AI521804

seq_documentation_block:
LOCUS       AI521804      599 bp    mRNA            EST       13-APR-1999
DEFINITION  ti82f04.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2138527 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI521804
VERSION     AI521804.1  GI:4435939
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948550.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 893   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2138527"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      138 a    178 c    181 g    101 t      1 others
ORIGIN


alignment_scores:
            Quality:  900.00            Length:    190
              Ratio:  4.891               Gaps:      0
 Percent Similarity:  96.842   Percent Identity: 92.632

alignment_block:
 US-09-304-121-2 x AI521804   ..

 Align seg 1/1  to: AI521804  from: 1  to: 599

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      20 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 69

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      70 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 119

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     120 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 169

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     170 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 219

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     220 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 269

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     270 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 319

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     320 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 369

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     370 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 419

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     420 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 469

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         :::||||||||||||||||||||||||||||||||||||||||||||||||
     470 CATGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 519

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||   ||||||||||||||:::      :::   :::
     520 TCTGTGGCGGGAGTG.GCCAGCCTTCGGCAGAGCATGCCCAGCACAGAAG 568

     184 ysValValAsnLysLeuIle 190
           |||:::::::::   :::
     569 TCGTCACAAGCTCATCAGTN 588

seq_name: gb_est35:AI863994
```

```
seq_documentation_block:
LOCUS       AI863994      522 bp    mRNA            EST       30-AUG-1999
DEFINITION  wj54e02.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2406650 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI863994
VERSION     AI863994.1  GI:5528025
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3136680.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2406650"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      120 a    154 c    164 g     84 t
ORIGIN


alignment_scores:
            Quality:  876.00             Length:    171
              Ratio:  5.214                Gaps:      1
 Percent Similarity:  98.246   Percent Identity: 98.246

alignment_block:
 US-09-304-121-2 x AI863994    ..

 Align seg 1/1  to: AI863994  from: 1  to: 522          .

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      11 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 60

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      61 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 110

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     111 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 160

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 210

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 260

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 310

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     311 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 360

     117 gLysValThr.SerValSerThrLeuLysSerGluAspIleLysIleArg 133
         ||||||||||    ||||||||||||||||||||||||||||||||||||
     361 GAAAGTGACCATGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGC 410

     134 GlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLy 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     411 CAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAA 460

     150 sGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluA 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     461 GCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGC 510

     167 laLeuTrpArg 170
         |||||||||
     511 CTCTGTGGCGG 521

seq_name: gb_est17:AA619537

seq_documentation_block:
LOCUS       AA619537      580 bp    mRNA            EST       09-OCT-1997
DEFINITION  vo84h01.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1065841 5' similar to gb:X61753 M.musculus mRNA for heat
            shock transcription factor 1 (MOUSE);, mRNA sequence.
ACCESSION   AA619537
VERSION     AA619537.1  GI:2523413
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1517425.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:588201
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
```

```
    source          1. .580
                    /organism="Mus musculus"
                    /strain="C3H"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:1065841"
                    /clone_lib="Barstead mouse myotubes MPLRB5"
                    /cell_line="C2C12"
                    /lab_host="DH10B"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                    3']; double-stranded cDNA was ligated to Eco RI adaptors
                    [AATTCGGATCCTTG], digested with Not I and cloned into the
                    Not I and Eco RI sites of the modified pT7T3 vector.
                    Library constructed by Bob Barstead. The C2C12 cell line
                    (available from ATCC, catalog # CRL-1772) differentiates
                    rapidly, forming contractile myotubes and producing
                    characteristic muscle proteins. "
BASE COUNT      155 a    141 c    169 g    115 t
ORIGIN


alignment_scores:
           Quality:  853.00            Length:    200
             Ratio:  4.466               Gaps:      2
 Percent Similarity:  95.500   Percent Identity: 92.000

alignment_block:
 US-09-304-121-2 x AA619537   ..

 Align seg 1/1  to: AA619537  from: 1  to: 580

      27 SerAspProAspThrAspAlaLeuIleCysTrp...SerProSerGlyAs 42
         |||||||||                     |||   :::||||||||||||
       4 TCGGATCCT.....................TGGCTGGAGCCGAGTGGGAA 32

      42 nSerPheHisValPheAspGlnGlyGlnPheAlaLysGluValLeuProL 59
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      33 CAGCTTCCACGTGTTTGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCA 82

      59 ysTyrPheLysHisAsnAsnMetAlaSerPheValArgGlnLeuAsnMet 75
         ||||||||||||||||||||||||||||||||||   |||||||||
      83 AGTACTTCAAGCACAACAACATGGCTAGCTTCGTGCGGCA.CTCAACATG 131

      76 TyrGlyPheArgLysValValHisIleGluGlnGlyGlyLeuValLysPr 92
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     132 TATGGCTTCCGAAAAGTAGTCCACATTGAGCAGGGTGGCCTGGTCAAGCC 181

      92 oGluArgAspAspThrGluPheGlnHisProCysPheLeuArgGlyGlnG 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     182 TGAGAGAGATGACACCGAGTTCCAGCATCCTTGTTTCTTGCGTGGACAGG 231

     109 luGlnLeuLeuGluAsnIleLysArgLysValThrSerValSerThrLeu 125
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     232 AACAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGCGTGTCCACCCTG 281

     126 LysSerGluAspIleLysIleArgGlnAspSerValThrLysLeuLeuTh 142
         ||||||||||||||||||||||||||||||||||||||:::|||||||||
     282 AAGAGTGAGGACATAAAAATACGCCAGGACAGTGTCACCCGGCTGTTGAC 331

     142 rAspValGlnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeuL 159
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     332 AGATGTGCAGCTGATGAAGGGGAAACAGGAGTGTATGGACTCCAAGCTCC 381

     159 euAlaMetLysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeu 175
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     382 TGGCCATGAAGCACGAGAACGAGGCCCTGTGGCGGGAGGTGGCCAGCCTT 431

     176 ArgGlnLysHisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPh 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     432 CGGCAGAAGCATGCCCAGCAGCAAAAAGTTGTCAACAAGCTCATTCAGTT 481

     192 eLeuIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysI 209
         ||||||||||||||||||||||||||||||||||||||||||||||||   ||||
     482 CCTGATCTCACTGGTGCAGTCGAACCGGATCCTGGGGGTGAAGAG.AAGA 530

     209 leProLeuMetLeuAsnAspSerGlySerAlaHisSerMetProLysTyr 225
         |||||||||||||||:::|||      |||||||:::|||:::|||||||||
     531 TCCCTCTGATGTTGAGTGACAGAAA.TCAGCACAATCTGTGCCCAAGTAT 579

seq_name: gb_est43:AW169960

seq_documentation_block:
LOCUS       AW169960      524 bp    mRNA            EST       12-NOV-1999
DEFINITION  xj35b05.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2659185 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1
            (HUMAN);, mRNA sequence.
ACCESSION   AW169960
VERSION     AW169960.1  GI:6401485
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 16, 1998 this sequence version replaced gi:2961802.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2659185"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GC1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      116 a    165 c    159 g     84 t
ORIGIN


alignment_scores:
           Quality:  846.00            Length:    161
             Ratio:  5.255               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-304-121-2 x AW169960   ..

 Align seg 1/1  to: AW169960  from: 1  to: 524

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      38 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 87
```

```
     17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     88 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 137

     34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    138 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 187

     51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    188 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 237

     67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    238 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 287

     84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    288 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 337

    101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    338 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 387

    117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    388 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 437

    134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    438 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 487

    151 GlnGluCysMetAspSerLysLeuLeuAlaMet 161
        |||||||||||||||||||||||||||||||
    488 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATG 520

seq_name: gb_est37:AI934773

seq_documentation_block:
LOCUS       AI934773      497 bp    mRNA            EST       02-SEP-1999
DEFINITION  wp89c05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2468936 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI934773
VERSION     AI934773.1  GI:5673643
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137023.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2468936"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      110 a    157 c    152 g     78 t
ORIGIN


alignment_scores:
            Quality:  804.00              Length:    153
              Ratio:  5.255                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.346

alignment_block:
 US-09-304-121-2 x AI934773   ..

 Align seg 1/1  to: AI934773  from: 1  to: 497

      1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     39 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 88

     17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     89 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 138

     34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    139 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 188

     51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    189 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 238

     67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    239 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 288

     84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    289 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 338

    101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    339 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 388

    117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    389 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 438

    134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    439 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 488

    151 GlnGluCys 153
        :::|||||||
    489 CACGAGTGC 497
```

THIS PAGE BLANK (USPTO)

OM of: US-09-304-121-2 to: N_Geneseq_36:* out_format : pfs

Date: Mar 7, 2000 1:09 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09304121/runat_06032000_151317_15740/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09304121 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-09-304-121-2
Query length: 529
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 72.590000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
N_Geneseq_36:Q13241	+	2729.00	2846.83	5.7e-151	2156	! Human HSF cDNA sequence. DNA e
N_Geneseq_36:V32958	+	2729.00	2846.83	5.7e-151	2156	! Human wild-type heat shock tra
N_Geneseq_36:Q25713	+	2689.00	2805.02	1.2e-148	2156	! Sequence of human Heat Shock F
N_Geneseq_36:Q13239	+	709.50	733.60	2.9e-33	2781	! HSF cDNA sequence. DNA encodin
N_Geneseq_36:Q25712	+	705.50	729.42	5.0e-33	2781	! Sequence of Drosophila heat sh
N_Geneseq_36:T84949	+	336.00	361.08	1.6e-12	350	! Human prostate protein HPA38 cD
N_Geneseq_36:Q23000	+	174.00	180.81	0.0180	1239	! Sequence encoding rye-grass pc
N_Geneseq_36:V02998	+	170.50	169.81	0.0739	2898	! Mouse neural Mena+ cDNA. Detec
N_Geneseq_36:T94101	+	160.00	133.63	7.66	53526	! Human PKD1 gene. Human polycy
N_Geneseq_36:T18551	+	160.00	133.62	7.66	53577	! Human polycystic kidney disea
N_Geneseq_36:T94108	+	160.00	133.62	7.66	53577	! Human PKD1 locus between chrc
N_Geneseq_36:Q53866	+	155.00	150.20	0.9137	4297	! Ge protein-encoding cDNA. Ge p
N_Geneseq_36:V05287	+	155.00	129.10	13.69	49377	! The soraphen biosynthesis gen
N_Geneseq_36:N90357	+	148.00	147.27	1.33	2589	! Sequence of a genetic construc
N_Geneseq_36:T06769	+	146.50	124.94	23.35	28598	! Sorangium cellulosum soraphen
N_Geneseq_36:T89956	+	146.50	124.83	23.68	28958	! Sorangium cellulosum soraphen
N_Geneseq_36:V22682	-	145.00	145.48	1.67	2214	! New DNA sequence isolated from
N_Geneseq_36:V22683	-	145.00	141.95	2.63	3331	! New DNA sequence isolated from
N_Geneseq_36:Q56747	+	143.50	148.03	1.21	1376	! Ryegrass Lol pV allergen. Pept
N_Geneseq_36:X32290	+	141.50	140.97	2.99	2444	! Chimeric transcription activat
N_Geneseq_36:T42902	+	141.50	129.66	12.74	9045	! DNA sequence which regulates e
N_Geneseq_36:N40080	+	141.50	129.66	12.74	9047	! Sequence complementary to the
N_Geneseq_36:Q03259	+	140.50	133.92	7.38	4897	! Pseudorabies virus glycoprotei
N_Geneseq_36:Q81790	+	140.50	128.70	14.41	8957	! HTLV-II virus NRA strain genom
N_Geneseq_36:Q73500	-	139.00	127.65	16.49	8438	! DNA encoding Pseudorabies viru
N_Geneseq_36:T80043	+	139.00	114.01	94.83	40875	! Insert from cosmid 109. New n
N_Geneseq_36:Q38229	+	137.50	142.73	2.38	1229	! Sequence of rye grass pollen c
N_Geneseq_36:Q85932	+	137.50	142.73	2.38	1229	! cDNA encoding Lol pV (clone 12
N_Geneseq_36:Q03324	-	137.00	127.72	16.33	6567	! Eimeria tenella genomic DNA en
N_Geneseq_36:V99229	+	136.50	144.14	1.99	925	! DNA encoding an active acyltran
N_Geneseq_36:T70123	+	136.50	131.25	10.38	4108	! DNA encoding beta-xylosidase (
N_Geneseq_36:Q28398	-	136.50	128.06	15.64	5946	! Rat nestin gene. Diagnosis of
N_Geneseq_36:Q70447	-	136.50	122.56	31.68	11236	! Rat nestin gene - its product
N_Geneseq_36:Q05226	+	136.00	138.33	4.19	1706	! Primer binding to HTLV-1 genom
N_Geneseq_36:V71087	+	135.50	134.88	6.52	2394	! PKB-green fluorescent protein
N_Geneseq_36:V32592	+	135.00	133.43	7.86	2665	! Schwannomin-binding protein cD
N_Geneseq_36:Q03317	-	133.00	141.07	2.95	864	! cDNA of Eimeria tenella oocyte
N_Geneseq_36:N60488	-	133.00	141.07	2.95	864	! Eimeria tenella 5401 sporozoite
N_Geneseq_36:T84934	+	133.00	128.64	14.53	3642	! Human prostate protein HPA32 c
N_Geneseq_36:Q58624	+	133.00	125.43	21.92	5278	! Mammalian son of sevenless gen
N_Geneseq_36:V56447	+	133.00	121.28	37.32	8532	! Human APC cDNA. Adenomatous po
N_Geneseq_36:Q27234	+	133.00	120.25	42.57	9606	! Encodes APC gene in familia
N_Geneseq_36:Q70633	+	133.00	120.25	42.57	9606	! Adenomatous polyposis coli
N_Geneseq_36:T95538	+	133.00	120.25	42.57	9606	! Human adenomatous Polyposis
N_Geneseq_36:T96153	+	133.00	120.25	42.57	9606	! Human adenomatous Polyposis

seq_name: N_Geneseq_36:Q13241

seq_documentation_block:
ID Q13241 standard; cDNA; 2156 BP.
AC Q13241;
DT 29-OCT-1991 (first entry)
DE Human HSF cDNA sequence.
KW Heat shock factor; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 160. .1750
FT /*tag= a
FT polya_site 2137. .2156
FT /*tag= b
PN US7617901-A.
PD 16-JUL-1991.
PF 26-NOV-1990; 617901.
PR 26-NOV-1990; US-617901.
PA (USSH) NAT INST OF HEALTH.
PI Wu C, Clos J, Westwood JT, Rabindran S;
DR WPI; 91-252343/34.
DR P-PSDB; R13503.
PT DNA encoding Drosophila and human heat shock factor proteins -
PT used for developing prods. for studying stress and disease states
PT in living systems.
PS Disclosure; Fig 13; 68pp; English.
CC The sequence encodes human heat shock factor protein and was
CC obtained by using short stretches of homologous sequences between
CC Drosophila and yeast HSFs as primers for the polymerase chain
CC reaction. The human HSF DNA was obtd. by screening human B cell
CC cDNA libraries with the amplified sequence. The HSF sequence can be
CC used to identify the HSF genes in other organisms and also for the
CC detection of stress or a diseased state in living systems. The gene
CC can be used to increase expression of other gene prods. by co-
CC transfecting the HSF gene together with other genes linked to heat
CC shock elements. It can be linked to a tissue-general or tissue
CC specific promoter and introduced into transgenic mice as a tool for
CC eliciting increased or chronic stress response conditions as a
CC model for how tissues respond to chronic stress conditions such as
CC those caused by viral infection, chemical or mechanical stress.
CC There is no stop codon upstream of the designated AUG start codon
CC so these sequences and the encoded amino acids, including thosee
CC further upstream not yet sequenced may form part of the natural HSF
CC protein. See also Q13240 and Q13241.
SQ Sequence 2156 BP; 435 A; 743 C; 624 G; 354 T;

alignment_scores:
Quality: 2729.00 Length: 529
Ratio: 5.159 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-304-121-2 x Q13241 ..

Align seg 1/1 to: Q13241 from: 1 to: 2156

```
      1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    161 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 210

     17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    211 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 260

     34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    261 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 310
```

```
 51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
311 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

 67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 410

 84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
411 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 460

101 HisProCysPheLeuArgGlyGlnGluGlnLeuGluAsnIleLysAr 117
    ||||||||||||||||||||||||||||||||||||||||||||||||
461 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 510

117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
511 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 560

134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
561 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 610

151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
611 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 710

184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
711 AAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAAC 760

201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
761 CGGATCCTGGGGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTGG 810

217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
811 CTCAGCACATTCCATGCCCAAGTATAGCCGGCAGTTCTCCCTGGAGCACG 860

234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
861 TCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTCC 910

251 SerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAs 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
911 AGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

267 pIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerI 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCA 1010

284 leAspGluArgProLeuSerSerSerProLeuValArgValLysGluGlu 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1011 TAGACGAGAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAG 1060

301 ProProSerProProGlnSerProArgValGluGluAlaSerProGlyAr 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1061 CCCCCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGCG 1110

317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1111 CCCATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCCA 1160

334 leLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAsp 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1161 TCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGAC 1210

351 AlaArgGlyHisThrAspThrGluGlyArgProProSerProProProTh 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1211 GCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

367 rSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGluL 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGC 1310

384 euSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThr 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1311 TCAGTGACCACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACC 1360

401 MetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLe 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1361 ATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCT 1410

417 uPheSerProSerValThrValProAspMetSerLeuProAspLeuAspS 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1411 GTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACA 1460

434 erSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProArg 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1461 GCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGG 1510

451 ProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValHi 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1511 CCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

467 sTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThrG 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCG 1610

484 lySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPhe 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1611 GGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTC 1660

501 SerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGl 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1661 TCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGG 1710

517 ySerGluProProLysAlaLysAspProThrValSer 529
    |||||||||||||||||||||||||||||||||||||
1711 CTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 1747
```

seq_name: N_Geneseq_36:V32958

seq_documentation_block:

```
ID   V32958 standard; DNA; 2156 BP.
AC   V32958;
DT   26-OCT-1998  (first entry)
DE   Human wild-type heat shock transcription factor 1 (HSF1) DNA.
KW   Human wild-type heat shock transcription factor 1; HSF1; ischaemia;
KW   heat shock protein; hsp; UV-B light; sepsis; hyperthermia;
KW   oxidative stress; anti-tumour agent; cancer cell; cytotoxic agent; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             161. .1750
FT                   /*tag=  a
FT                   /product= "HSF1"
PN   WO9831803-A1.
PD   23-JUL-1998.
PF   21-JAN-1998; U01038.
PR   19-AUG-1997; US-914646.
PR   21-JAN-1997; US-035662.
PA   (UYMI-) UNIV MIAMI.
PI   Voellmy RW;
DR   WPI; 98-414102/35.
DR   P-PSDB; W49093.
PT   Method for modulating synthesis of heat-shock protein - by
PT   administering mutant heat shock transcription factors, used, e.g to
PT   protect cells against chemotherapy
PS   Disclosure; Fig 1A-1C; 84pp; English.
CC   The present sequence represents the wild-type heat shock transcription
```

```
CC   factor 1 (HSF1) DNA.  The invention provides a method for modulating
CC   the expression of endogenous heat shock protein (hsp) genes in
CC   eukaryotic cells.  The method involves introducing a mutated HSF1
CC   or a gene encoding the mutated HSF1 into cells, with the result
CC   that hsp synthesis in the recipient cells is altered.  Mutation of
CC   a regulatory region, spanning residues 180 to 315, in the HSF1
CC   sequence resulted in hsp synthesis activation in the absence of
CC   stress.  This positive acting mutant HSF1 is claimed to induce a
CC   protected state in the cell.  Mutation of a second region, spanning
CC   residues 277 to 529, in the HSF1 sequence resulted in hsp synthesis
CC   inhibition induced by stress in the presence of stress.  This negative
CC   acting mutated HSF1 is claimed to induce a sensitised state in a cell.
CC   The mutant HSF1s are claimed to be useful for protecting cells against
CC   damage caused by therapeutics, UV-B light, sepsis, hyperthermia,
CC   oxidative stress and ischaemia, particularly to increase resistance of
CC   normal cells to anti-tumour agents, or to increase immunogenicity of
CC   cancer cells.  The mutant HSF1s  are active in absence of stress,
CC   unlike wild-type HSF1, even when over expressed, and eliminate the need
CC   for cytotoxic agents for regulating the heat-shock system.
SQ   Sequence   2156 BP;    435 A;    739 C;    628 G;    354 T;


alignment_scores:
            Quality: 2729.00              Length:    529
              Ratio:  5.159                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-304-121-2 x V32958   ..

 Align seg 1/1  to: V32958  from: 1  to: 2156

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     161 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 210

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 260

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 310

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     311 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 410

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     411 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 460

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     461 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 510

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     511 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 560

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     561 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 610

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     611 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 710

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     711 AAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAAC 760

     201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     761 CGGATCCTGGGGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTGG 810

     217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     811 CTCAGCACATTCCATGCCCAAGTATAGCCGGCAGTTCTCCCTGGAGCACG 860

     234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     861 TCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTCC 910

     251 SerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     911 AGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

     267 pIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerI 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCA 1010

     284 leAspGluArgProLeuSerSerSerProLeuValArgValLysGluGlu 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1011 TAGACGAGAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAG 1060

     301 ProProSerProProGlnSerProArgValGluGluAlaSerProGlyAr 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1061 CCCCCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGCG 1110

     317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1111 CCCATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCCA 1160

     334 leLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAsp 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1161 TCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGAC 1210

     351 AlaArgGlyHisThrAspThrGluGlyArgProProSerProProProTh 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1211 GCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

     367 rSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGluL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGC 1310

     384 euSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThr 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1311 TCAGTGACCACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACC 1360

     401 MetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1361 ATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCT 1410

     417 uPheSerProSerValThrValProAspMetSerLeuProAspLeuAspS 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1411 GTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACA 1460

     434 erSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProArg 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1461 GCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGG 1510

     451 ProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValHi 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1511 CCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

     467 sTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThrG 484
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCG 1610

     484 lySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPhe 500
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1611 GGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTC 1660

     501 SerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGl 517
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1661 TCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGG 1710

     517 ySerGluProProLysAlaLysAspProThrValSer 529
         |||||||||||||||||||||||||||||||||||||
    1711 CTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 1747

seq_name: N_Geneseq_36:Q25713

seq_documentation_block:
ID   Q25713 standard; cDNA; 2156 BP.
AC   Q25713;
DT   28-DEC-1992  (first entry)
DE   Sequence of human Heat Shock Factor (HSF) cDNA
KW   Heat shock factor; stress condition; assay; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             160. .1750
FT                   /*tag= a
PN   WO9209617-A.
PD   11-JUN-1992.
PF   22-NOV-1991; U08592.
PR   26-NOV-1990; US-617910.
PA   (USDC ) US DEPT OF COMMERCE.
PI   Clos J, Rabindran S, Westwood JT, Wu C;
DR   WPI; 92-217013/26.
DR   P-PSDB; R24948.
PT   DNA fragment encoding Drosophila or human heat shock factor
PT   protein - and use of corresp. monoclonal antibodies for
PT   diagnosing abnormal stress conditions in cells
PS   Claim 5; Figure 13; 75pp; English.
CC   The cloning of human heat shock factor HuHSF) was achieved by using
CC   short stretches of homologous sequences between Drosophila and
CC   yeast heat shock factors as primers in the polymerase chain
CC   reaction (PCR) (Q25714,Q25715). The HuHSF length clone was obtained
CC   by screening human cDNA libraries with the amplified sequence. The
CC   HuHSF cDNA clone includes an open reading frame of 529 AAs with a
CC   calculated molecular weight of 58,000 (Q24713,R24948). The size of
CC   HuHSF as measured by SDS-polyacrylamide gel electrophoresis is
CC   60,000 which is in close agreement with the calculated size. The
CC   claims refer to Figure 12, rather than Figure 13, but this would
CC   appear to be an error in the claims.
SQ   Sequence   2156 BP;    435 A;    739 C;    628 G;    354 T;


alignment_scores:
            Quality: 2689.00             Length:    530
              Ratio:  5.093               Gaps:      1
 Percent Similarity:  99.623   Percent Identity: 99.623

alignment_block:
 US-09-304-121-2 x Q25713   ..

 Align seg 1/1  to: Q25713  from: 1  to: 2156

       1 MetAspLeuProValGlyProGlyAlaAla.GlyProSerAsnValProA 17
         |||||||||||||||||||||||||||||| ||||||||||||||||||||
     161 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCCAGCAACGTCCCGG 210

      17 laPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAla 33
         |||||||||||||||||||||||   |||||||||||||||||||||||
     211 CCTTCCTGACCAAGCTGTGGACCTC.GTGAGCGACCCGGACACCGACGCG 259

      34 LeuIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGl 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  .  260 CTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGG 309

      50 yGlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetA 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    310 CCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGG 359

      67 laSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHis 83
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    360 CCAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCAC 409

      84 IleGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGl 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    410 ATCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCA 459

     100 nHisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysA 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    460 GCACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGA 509

     117 rgLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArg 133
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    510 GGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGC 559

     134 GlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLy 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    560 CAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAA 609

     150 sGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluA 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    610 GCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGG 659

     167 laLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGln 183
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    660 CTCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAG 709

     184 LysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAs 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    710 AAAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAA 759

     200 nArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerG 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    760 CCGGATCCTGGGGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTG 809

     217 lySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHis 233
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    810 GCTCAGCACATTCCATGCCCAAGTATAGCCGGCAGTTCTCCCTGGAGCAC 859

     234 ValHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSe 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    860 GTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTC 909

     250 rSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerA 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    910 CAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCG 959

     267 spIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySer 283
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    960 ACATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGC 1009

     284 IleAspGluArgProLeuSerSerSerProLeuValArgValLysGluGl 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1010 ATAGACGAGAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGA 1059

     300 uProProSerProProGlnSerProArgValGluGluAlaSerProGlyA 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1060 GCCCCCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGC 1109

     317 rgProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSer 333
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1110 GCCCATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCC 1159

     334 IleLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAs 350
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1160 ATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGA 1209
```

```
     350 pAlaArgGlyHisThrAspThrGluGlyArgProProSerProProProT 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1210 CGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCA 1259

     367 hrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGlu 383
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1260 CCTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAG 1309

     384 LeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnTh 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1310 CTCAGTGACCACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGAC 1359

     400 rMetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspL 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1360 CATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACC 1409

     417 euPheSerProSerValThrValProAspMetSerLeuProAspLeuAsp 433
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1410 TGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGAC 1459

     434 SerSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProAr 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1460 AGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAG 1509

     450 gProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValH 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1510 GCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGC 1559

     467 isTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThr 483
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1560 ACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACC 1609

     484 GlySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPh 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1610 GGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTT 1659

     500 eSerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrG 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1660 CTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAG 1709

     517 lySerGluProProLysAlaLysAspProThrValSer 529
         ||||||||||||||||||||||||||||||||||||||
    1710 GCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 1747

seq_name: N_Geneseq_36:Q13239

seq_documentation_block:
ID   Q13239 standard; cDNA; 2781 BP.
AC   Q13239;
DT   29-OCT-1991  (first entry)
DE   HSF cDNA sequence.
KW   Heat shock factor; ss.
OS   Drosophila.
FH   Key             Location/Qualifiers
FT   cds             229. .2304
FT                   /*tag=  a
FT   polya_signal    2722. .2727
FT                   /*tag=  b
FT   polya_site      1757. .1781
FT                   /*tag=  c
PN   US7617901-A.
PD   16-JUL-1991.
PF   26-NOV-1990; 617901.
PR   26-NOV-1990; US-617901.
PA   (USSH ) NAT INST OF HEALTH.
PI   Wu C, Clos J, Westwood JT, Rabindran S;
DR   WPI; 91-252343/34.
DR   P-PSDB; R13502.
PT   DNA encoding Drosophila and human heat shock factor proteins -
PT   used for developing prods. for studying stress and disease states
PT   in living systems.
PS   Disclosure; Fig 2; 68pp; English.
CC   The sequence encodes Drosophila heat shock factor protein and was
CC   obtained by screening a Drosophila genomic library with oligo-
CC   nucleotide probes (Q13237, Q13238) based on the HSF amino acid
CC   sequence. The HSF sequence can be used to identify the HSF genes in
CC   other organisms and also for the detection of stress or a diseased
CC   state in living systems. The gene can be used to increase
CC   expression of other gene prods. by cotransfecting the HSF gene
CC   together with other genes linked to heat shock elements. It can be
CC   linked to a tissue-general or tissue-specific promoter and
CC   introduced into transgenic mice as a tool for eliciting increased
CC   or chronic stress response conditions as a model for how tissues
CC   respond to chronic stress conditions such as those caused by viral
CC   infection, chemical or mechanical stress. See also Q13240 and
CC   Q13241.
SQ   Sequence   2781 BP;    831 A;    631 C;    690 G;    629 T;


alignment_scores:
           Quality:  709.50              Length:    662
             Ratio:  2.027                Gaps:     22
 Percent Similarity:  52.870   Percent Identity: 30.514

alignment_block:
 US-09-304-121-2 x Q13239   ..

 Align seg 1/1  to: Q13239  from: 1  to: 2781

       6 GlyProGlyAlaAlaGlyProSerAsnValProAlaPheLeuThrLysLe 22
         |||   :::||||||      |||:::|||||||||||||||:::|||||
     337 GGAGACGCCGCGGCCATCGGAAGCGGGGTGCCGGCCTTTTTGGCCAAATT 386

      22 uTrpThrLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerP 39
         ||||   ||||||:::|||   ||||||:::   ||||||||||||:::
     387 GTGGCGCCTGGTGGACGATGCCGATACCAATCGCTTGATTTGCTGGACCA 436

      39 roSerGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGlu 55
           :::|||:::|||||   :::   :::|||:::|||||||||||||||
     437 AGGATGGCCAAAGTTTCGTTATTCAAAATCAAGCGCAATTTGCCAAGGAA 486

      56 ValLeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGl 72
         :::|||||      :::|||||||||||||||||||||||||:::|||||
     487 CTATTGCCACTAAACTACAAGCACAACAACATGGCCAGTTTCATAAGGCA 536

      72 nLeuAsnMetTyrGlyPheArgLysValValHisIleGluGlnGlyGlyL 89
         ||||||||||||||||||:::|||::::::   |||:::::|||||||
     537 ATTGAATATGTATGGATTCCACAAGATCACCTCTATTGACAATGGCGGAC 586

      89 euValLysProGluArgAspAspThrGluPheGlnHisProCysPheLeu 105
         ||   :::   :::||||||:.   ||||||:::||||||   |||
     587 TA...CGTTTTGATCGCGACGAGATTGAATTTTCGCACCCATTTTTTAAG 633

     106 ArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerVa 122
         |||::::::      ||||||:::::|||||||||||         ::
     634 CGCAACTCGCCTTTTCTACTTGACCAAATCAAAAGGAAA.........AT 674

     122 lSerThrLeuLysSerGluAspIleLys......IleArgGlnAspSerV 137
         :|||:::   |||:::   |||   |||      ::::::   :::::::
     675 ATCGAACAACAAAAATGGTGACGACAAAGGTGTCCTGAAGCCGGAGGCCA 724

     137 alThrLysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCys 153
         :::::|||:::||||||||||||:::::|||:::|||:::|||:::
     725 TGTCGAAGATTCTCACCGATGTGAAAGTCATGCGGGGTCGTCAGGACAAT 774

     154 MetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpAr 170
         :::|||||::::::   |||||||||:::||||||||:::|||||||||
     775 CTGGATTCGCGCTTCTCCGCCATGAAACAGGAGAACGAAGTGCTGTGGCG 824

     170 gGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValA 187
         ||||:::|||||||||||||||||||||||:::|||||:::::|||||
     825 CGAGATAGCCAGCCTGCGCCAAAAGCACGCTAAGCAGCAACAAATAGTCA 874

     187 snLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsnArg...Ile 202
         ||||||||||||||||||||||:::::|||||   :::|||   :::
     875 ACAAACTGATCCAGTTCCTCATTACCATTGTGCAACCGTCGCGCAACATG 924
```

```
 203 LeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySerAl 219
 925 TCTGGCGTGAAACGCCATGTGCAGCTGATGATGAACAATACG........ 966

 219 aHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisValHisG 236
 967 .CCGGAAATTGATCGTGCACGGAGGACCAGTGAGACCGAGAGCGAGAGTG 1015

 236 lySerGlyPro.................................Tyr 240
1016 GCGGCGGACCGGTTATCCACGAGCTTAGGGAGGAGCTTGTTGATGAGGTG 1065

 241 SerAlaProSerProAla...TyrSerSerSerSerLeuTyrAlaProAs 256
1066 ATGAATCCATCACCGGCTGGCTACACCGCAGCCTCACATTATGACCAAGA 1115

 256 pAlaValAla..................................... 259
1116 GAGCGTCTCTCCGCCTGCCGTTGAGCGTCCGCGATGTAAGATGAGCATTA 1165

 259 ..... 259

1166 GCTGGCACAACGTCGATTATTCGAATCAGAGTGTGGAGGACTTGCTGCTC 1215

 260 ...SerSerGlyProIleIleSerAspIleThrGluLeuAlaProAlaSe 275
1216 CAGGGAAATGGAACCGCTGGCGGTAATATTGTAGTAGGCGGAGCCGCTTC 1265

 275 rProMetAlaSerProGlyGlySerIleAspGluArgPro......... 288
1266 TCCCATGGCCCAA.........AGTGTGAGTCAATCGCCGGCCCAACATG 1306

 289 ..........LeuSerSerSerProLeuValArgValLysGluGluPro 301
1307 ATGTCTACACAGTCACCGAGGCGCCCGATTCTCATGTCCAGGAGGTGCCA 1356

 302 ProSerProPro.....................GlnSerProArgVa 310
1357 AACAGTCCGCCTTATTACGAGGAGCAGAATGTGCTTACCACGCCCATGGT 1406

 310 lGluGlu..... 312
1407 GCGGGAGCAGGAGCAGCAGAAGCATCAGCAGCTTAAGGAGAACAACAAGC 1456

 312 ..... 312

1457 TACGACGACAGGCAGGAGATGTTATACTTGATGCTGGAGATATTCTCGTA 1506

 313 ......AlaSerPro...GlyArgProSerSerValAspThrLeuLeuSe 326
1507 GATAGTTCGTCGCCCAAGGCGCAACGGACAAGCATCCAGCATAGTACGCA 1556

 326 rProThrAlaLeuIleAspSerIleLeuArgGluSerGluProAlaProA 343
1557 ACCTGATGTGATGGTCCAGCCAATGATTATAAAGTCTGAGCCGGAGAACA 1606

 343 laSer......ValThrAlaLeuThrAspAlaArgGly............ 353
1607 GTTCCGGACTGATGGATCTAATGACTCCCGCGAACGATCTGTACAGTGTC 1656

 353 ..... 353

1657 AACTTCATCAGTGAGGATATGCCGACGGATATTTTTGAAGACGCTGTGCT 1706

 353 ..... 353

1707 TCCCGACGGCGTGGAAGAGGCAGCCAAACTGGACCAGCAGCAAAAATTTG 1756

 354 .....HisThrAspThrGluGlyArgProProSer.........ProPro 365
1757 GGCAATCGACAGTGAGCAGCGGCAAGTTTGCCAGCAACTTCGATGTGCCC 1806

 366 ProThrSerThr..... 369
1807 ACCAACAGTACGCTGCTGGATGCCAATCAGGCCTCGACATCGAAGGCAGC 1856

 370 .....ProGluLysCysLeuSerValAla..... 377
1857 GGCCAAGGCGCAAGCATCTGAGGAAGAGGGCATGGCTGTGGCCAAATACA 1906

 377 ..... 377

1907 GTGGCGCTGAGAACGGAAACAACCGCGATACCAACAACAGTCAACTCCTC 1956

 378 ...CysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSe 393
1957 AGGATGGCCTCAGTTGACGAACTCCACGGGCACTTGGAAAGCATGCAAGA 2006

 393 rAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValA 410
2007 TGAGTTGGAAACACTGAAGGATCTGCTGCGCGGCGATGGGGTGGCCATTG 2056

 410 spThrSerAlaLeuLeuAspLeuPheSerProSerValThrValProAsp 426
2057 ATCAGAACATGCTCATGGGTCTGTTTAACGACTCTGATCTAATGGACAAC 2106

 427 MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLe 443
2107 TATGGCCTATCGTTTCCCAATGACAGCATAAGCAGT.............. 2142

 443 uSerProGlnGluProProArgProProGluAlaGluAsnSerSerProA 460
2143 ..............................GAAAAGAAAGCACCC. 2157

 460 spSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPhe...Leu 475
2158 ..AGTGGCTCTGAACTGATTTCCTAT......CAGCCCATGTATGATCTG 2199

 476 LeuAspProGlySerValAspThrGlySerAsnAsp 487
2200 TCCGACATTTTGGACACGGACGATGGCAACAATGAC 2235

seq_name: N_Geneseq_36:Q25712

seq_documentation_block:
ID   Q25712 standard; cDNA; 2781 BP.
AC   Q25712;
DT   28-DEC-1992  (first entry)
DE   Sequence of Drosophila heat shock factor (HSF) cDNA.
KW   Heat shock factor; stress condition; assay; ss.
OS   Drosophila.
FH   Key             Location/Qualifiers
FT   cds             229. .2315
FT                   /*tag= a
FT   polya_signal    2723. .2728
FT                   /*tag= b
PN   WO9209617-A.
PD   11-JUN-1992.
PF   22-NOV-1991; U08592.
PR   26-NOV-1990; US-617910.
PA   (USDC ) US DEPT OF COMMERCE.
PI   Clos J, Rabindran S, Westwood JT, Wu C;
DR   WPI; 92-217013/26.
DR   P-PSDB; R24947.
PT   DNA fragment encoding Drosophila or human heat shock factor
PT   protein - and use of corresp. monoclonal antibodies for
PT   diagnosing abnormal stress conditions in cells
PS   Claim 3; Fig 2B; 75pp; English.
CC   Two 20-mer oligonucleotides with 32-fold degeneracy (Q25710,Q25711),
CC   based on the predicted nucleotide sequences of HSF peptide 27 and
CC   peptide 29 were used to probe a Drosophila genomic library.
CC   Initially two genomic DNA clones were identified which contained a
CC   common, 1800nt SalI-EcoRI fragment. This fragment, which hybridised
CC   with both oligo probes, was then used to isolate cDNA clones from a
CC   random-primed and an oligo dT-primed cDNA library. The 2.8 kb of HSF
```

```
CC   cDNA sequence reconstructed from six overlapping cDNA clones reveals
CC   a single open reading frame of 691 amino acids. The sequences of all
CC   six HSF tryptic peptides within the 691-amino acid open reading frame
CC   were located, and thus concluded that this reading frame encodes
CC   Drosophila HSF (Q25712,R24947). The molecular mass of Drosophila
CC   HSF, calculated from the deduced amino acid sequence is 77,300
CC   daltons, significantly lower than the apparent mass of 110,000
CC   daltons measured by SDS gel electrophoresis. Evidently, Drosophila
CC   HSF has an anomalous mobility on SDS gels. Fig 2B (Q25712) has a
CC   non-standard nt (D) at posn. 741.
SQ   Sequence   2781 BP;    830 A;    639 C;    682 G;    629 T;


alignment_scores:
           Quality:  705.50            Length:    665
             Ratio:  2.016               Gaps:     23
 Percent Similarity:  52.632   Percent Identity: 30.526

alignment_block:
 US-09-304-121-2 x Q25712   ..

 Align seg 1/1  to: Q25712  from: 1  to: 2781

       6 GlyProGlyAlaAlaGlyProSerAsnValProAlaPheLeuThrLysLe 22
         |||   :::||||||       |||:::||||||||||||||||:::|||||
     337 GGAGACGCCGCGGCCATCGGAAGCGGGGTGCCGGCCTTTTTGGCCAAATT 386

      22 uTrpThrLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerP 39
         ||||   ||||||:::|||   ||||||:::   ||||||||||||:::
     387 GTGGCGCCTGGTGGACGATGCCGATACCAATCGCTTGATTTGCTGGACCA 436

      39 roSerGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGlu 55
           :::|||:::||||||   :::   :::|||:::||||||||||||||||
     437 AGGATGGCCAAAGTTTCGTTATTCAAAATCAAGCGCAATTTGCCAAGGAA 486

      56 ValLeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGl 72
         :::||||||       :::||||||||||||||||||||||||:::||||||
     487 CTATTGCCACTAAACTACAAGCACAACAACATGGCCAGTTTCATAAGGCA 536

      72 nLeuAsnMetTyrGlyPheArgLysValValHisIleGluGlnGlyGlyL 89
         ||||||||||||||||||:::|||:::::   |||:::::|||||||
     537 ATTGAATATGTATGGATTCCACAAGATCACCTCTATTGACAATGGCGGAC 586

      89 euValLysProGluArgAspAspThrGluPheGlnHisProCysPheLeu 105
         ||   :::   :::||||||:::   ||||||:::||||||   |||
     587 TA...CGTTTTGATCGCGACGAGATTGAATTTTCGCACCCATTTTTTAAG 633

     106 ArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerVa 122
         |||::::::       ||||||:::::::||||||||||         ::
     634 CGCAACTCGCCTTTTCTACTTGACCAAATCAAAAGGAAA.....AT 674

     122 lSerThrLeuLysSerGluAspIleLys......IleArgGlnAspSerV 137
         :|||:::   |||:::   |||   |||      ::::::   :::::::
     675 ATCGAACAACAAAAATGGTGACGACAAAGGTGTCCTGAAGCCGGAGGCCA 724

     137 alThrLysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCys 153
         :::::|||:::||||||||||||:::::|||:::|||:::|||:::
     725 TGTCGAAGATTCTCACDGATGTGAAAGTCATGCGGGGTCGTCAGGACAAT 774

     154 MetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpAr 170
         :::||||||::::::   |||||||||:::||||||||||:::||||||||
     775 CTGGATTCGCGCTTCTCCGCCATGAAACAGGAGAACGAAGTGCTGTGGCG 824

     170 gGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValA 187
         ||||:::|||||||||||||||||||||||:::||||||::::::||||
     825 CGAGATAGCCAGCCTGCGCCAAAAGCACGCTAAGCAGCAACAAATAGTCA 874

     187 snLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsnArg...Ile 202
         ||||||||||||||||||||||::::::||||||   :::|||   :::
     875 ACAAACTGATCCAGTTCCTCATTACCATTGTGCAACCGTCGCGCAACATG 924

     203 LeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySerAl 219
           ||||||||||||   :::   |||||||:::|||::::::

     925 TCTGGCGTGAAACGCCATGTGCAGCTGATGATCAACAATACG..... 966

     219 aHisSerMetProLysTyrSerArg.....GlnPheSerLeuGluH 233
                 |||:::   :::|||             :::   |||
     967 .....CCGGAAATTGATCGTGCACGGACCACCAGTGAGACCGAGA 1006

     233 isValHisGlySerGlyPro..... 239
               |||:::||||||
    1007 GCGAGAGTGGCGGCGGACCGGTTATCCACGAGCTTAGGGAGGAGCTTCTT 1056

     240 ......TyrSerAlaProSerProAla...TyrSerSerSerSerLeuTy 253
                       ||||||||||||   |||:::::::::|||   ||
    1057 GATGAGGTGATGAATCCATCACCGGCTGGCTACACCGCAGCCTCACATTA 1106

     253 rAlaProAspAlaValAla..... 259
         |     ::::::|||:::
    1107 TGACCAAGAGAGCGTCTCTCCGCCTGCCGTTGAGCGTCCGCGATCTAACA 1156

     259 ..... 259

    1157 TGAGCATTAGCTCGCACAACGTCGATTATTCGAATCAGAGTGTGGAGGAC 1206

     260 ............SerSerGlyProIleIleSerAspIleThrGluLeuAl 272
                     ::::::|||         ::::::|||          ::
    1207 TTGCTGCTCCAGGGAAATGGAACCGCTGGCGGTAATATTCTAGTAGGCGG 1256

     272 aProAlaSerProMetAlaSerProGlyGlySerIleAspGluArgPro. 288
         :   ||||||||||||||||:::          |||:::::::::   |||
    1257 AGCCGCTTCTCCCATGGCCCAA.....AGTGTGAGTCAATCGCCGG 1297

     289 ....................LeuSerSerSerProLeuValArgValLys 298
                              :::::::::::|||      :::|||:::
    1298 CCCAACATGATGTCTACACAGTCACCGAGGCGCCCGATTCTCATGTCCAG 1347

     299 GluGluProProSerProPro..... 305
         |||   |||   |||||||||
    1348 GAGGTGCCAAACAGTCCGCCTTATTACGAGGAGCAGAATGTGCTTACCAC 1397

     306 .....GlnSerProArg..... 309
                       |||:::   |||
    1398 GCCCATGGTGCGGGAGCAGGAGCAGCAGAAGCGTCAGCAGCTTAAGGAGA 1447

     309 ..... 309

    1448 ACAACAAGCTACGACGACAGGCAGGAGATGTTATACTTGATGCTGGAGAT 1497

     310 ......ValGluGluAlaSerPro...GlyArgProSerSerValAspTh 323
               |||::::::::::||||||   ::::::   :::|||::::::
    1498 ATTCTCGTAGATAGTTCGTCGCCCAAGGCGCAACGGACAAGCATCCAGCA 1547

     323 rLeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluP 340
               :::|||   ::::::::::::   ::::::   :::||||||||
    1548 TAGTACGCAACCTGATGTGATGGTCCAGCCAATGATTATAAAGTCTGAGC 1597

     340 roAlaProAlaSer......ValThrAlaLeuThrAspAlaArgGly... 353
         ||      :::|||       :::      :::|||   |||:::
    1598 CGGAGAACAGTTCCGGACTGATGGATCTAATGACTCCCGCGAACGATCTG 1647

     353 ..... 353

    1648 TACAGTGTCAACTTCATCAGTGAGGATATGCCGACGGATATTTTTGAAGA 1697

     353 ..... 353

    1698 CGCTCTGCTTCCCGACGGCGTGGAAGAGGCAGCCAAACTGGACCAGCAGC 1747

     354 ..............HisThrAspThrGluGlyArgProProSer..... 363
                         |||   ::::::|||:::      |||
    1748 AAAAATTTGGGCAATCGACAGTGAGCAGCGGCAAGTTTGCCAGCAACTTC 1797

     364 ...ProProProThrSerThr..... 369
               |||   :::||||||
    1798 GATGTGCCCACCAACAGTACGCTGCTGGATGCCAATCAGGCCTCGACATC 1847
```

```
     370 ..........................ProGluLysCysLeuSerValA 377
                                    |||:::   ::::::|||| 
    1848 GAAGGCAGCGGCCAAGGCGCAAGCATCTGAGGAAGAGGGCATGGCTGTGG 1897

     377 la..................................................... 377
         ||
    1898 CCAAATACAGTGGCGCTGAGAACGGAAACAACCGCGATACCAACAACAGT 1947

     378 ............CysLeuAspLysAsnGluLeuSerAspHisLeuAspAl 390
                      :::   :::|||||||      ||||||:::::
    1948 CAACTCCTCAGGATGGCCTCAGTTGACGAACTCCACGGGCACTTGGAAAG 1997

     390 aMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyP 407
         :|||::::::::::|||:::::::|||:::   :::|||   :::   |||
    1998 CATGCAAGATGAGTTGGAAACACTGAAGGATCTGCTGCGCGGCGATGGGG 2047

     407 heSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerValThr 423
           ::::::|||   :::   |||:::   ||||||:::   |||
    2048 TGGCCATTGATCAGAACATGCTCATGGGTCTGTTTAACGACTCTGATCTA 2097

     424 ValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGl 440
         :::   :::   :::|||         :::::::|||:::::::|||
    2098 ATGGACAACTATGGCCTATCGTTTCCCAATGACAGCATAAGCAGT..... 2142

     440 nGluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnS 457
                                                 |||:::: 
    2143 .....................................GAAAAGA 2149

     457 erSerProAspSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeu 473
         :::::|||   ||||||:::::::|||:::   |||      ||||||::: 
    2150 AAGCACCC...AGTGGCTCTGAACTGATTTCCTAT......CAGCCCATG 2190

     474 Phe...LeuLeuAspProGlySerValAspThrGlySerAsnAsp 487
         :::   |||   |||      :::::::|||   |||:::|||||| 
    2191 TATGATCTGTCCGACATTTTGGACACGGACGATGGCAACAATGAC 2235
```

seq_name: N_Geneseq_36:T84949

seq_documentation_block:
ID T84949 standard; cDNA; 350 BP.
AC T84949;
DT 27-APR-1998 (first entry)
DE Human prostate protein HPA38 cDNA.
KW Prostate cancer; immunotherapy; therapy; immunodiagnosis; diagnosis;
KW vaccine; human; HPA38; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 3. .350
FT /*tag= a
PN WO9733909-A2.
PD 18-SEP-1997.
PF 14-MAR-1997; U04192.
PR 11-APR-1996; US-633840.
PR 15-MAR-1996; US-616745.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Reed SG, Twardzik DR;
DR WPI; 97-470816/43.
DR P-PSDB; W27306.
PT Immunogenic portions of prostate proteins - useful to develop
PT products to detect, monitor, treat or inhibit development of
PT prostate cancer
PS Claim 28; Page 69; 84pp; English.
CC This cDNA sequence includes a coding region for human prostate
CC protein HPA38 (see W27306), an immunogenic portion of which can be
CC used in a claimed pharmaceutical composition for the treatment of
CC prostate cancer, in a claimed vaccine for treatment of prostate
CC cancer, or to raise claimed antibodies suitable for use in
CC diagnosis or monitoring the progression of prostate cancer. HPA38
CC cDNA was isolated from a human prostate adenocarcinoma cell line
CC LnCap.fgc (ATCC CRL-1740) cell cDNA library by expression screening
CC with human prostatitis sera. DNA sequences (see T84927-52) for 17
CC HPA proteins (see W23312-23 and W27303-07) are claimed and can be
CC used to produce recombinant HPA polypeptides in host cells
CC (particularly E. coli, yeast and mammalian cell lines) and to
CC design primers and probes for use in claimed methods of detecting
CC prostate cancer.
SQ Sequence 350 BP; 102 A; 77 C; 82 G; 89 T;

alignment_scores:
Quality: 336.00 Length: 93
Ratio: 4.148 Gaps: 2
Percent Similarity: 87.097 Percent Identity: 69.892

alignment_block:
US-09-304-121-2 x T84949 ..

Align seg 1/1 to: T84949 from: 1 to: 350

```
      13 SerAsnValProAlaPheLeuThrLysLeuTrpThrLeuValSerAspPr 29
         ||||||||||||||||||||:::||||||||||||||||||:::::: 
      72 TCGAACGTGCCGGCTTTCCTCAGCAAGCTGTGGACGCTTGTGGAGGAAAC 121

      29 oAspThrAspAlaLeuIleCysTrpSerProSerGlyAsnSerPheHisV 46
            |||:::   :::|||   ||||||   :::|||:::||||||   |
     122 CCACACTAACGAGTTCATCACCTGGAGCCAGAATGGCCAAAGTTTTCTGG 171

      46 alPheAspGlnGlyGlnPheAlaLysGluValLeuProLysTyrPheLys 62
         ||:::|||:::   :::|||||||||||:::|||||||||||||||||| 
     172 TCTTGGATGAGCAACGATTTGCAAAAGAAATTCTTCCCAAATATTTCAAG 221

      63 HisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheAr 79
         |||||||||||||||||||||||||||||||||||||||||||||||||| 
     222 CACAATAATATGGCAAGCTTTGTGAGGCAACTGAATATGTATGGTTTCCG 271

      79 gLysValValHisIleGluGlnGlyGlyLeuValLysProGluArgAsp. 95
         |||||||:::|||||||:::   :::|||:::||||||   |||||||| 
     272 TAAAGTAATACATATCGAC...TCTGGAATTGTTAAGCAAGAAAGAGATG 318

      96 ..AspThrGluPheGlnHisProCysPhe 104
            :::|||||||||||||||   |||
     319 GTCCTGTAGAATTTCAGCATCCTTACTTC 347
```

seq_name: N_Geneseq_36:Q23000

seq_documentation_block:
ID Q23000 standard; cDNA; 1239 BP.
AC Q23000;
DT 22-JUL-1992 (first entry)
DE Sequence encoding rye-grass pollen allergens encoded by EcoR1 insert
DE lambda-12R.
KW Rye grass pollinosis; diagnosis; therapy; ss.
OS Lolium perenne.
FH Key Location/Qualifiers
FT signal_peptide 41. .114
FT /*tag= a
FT mat_peptide 115. .966
FT /*tag= b
PN WO9203550-A.
PD 05-MAR-1992.
PF 16-AUG-1991; AU0369.
PR 17-AUG-1990; AU-001823.
PA (UYME-) UNIV MELBOURNE.
PI Singh MB, Hough T, Knox RB, Avjioglu A;
DR WPI; 92-096894/12.
DR P-PSDB; R22248.
PT New nucleic acid sequences coding rye-grass pollen allergens -
PT esp. Lol pIa and Lol pIb and their fragments, for diagnosing and
PT detecting rye-grass pollinosis
PS Disclosure; Fig3B; 81pp; English.
CC The inventors claim a sequence encoding the rye grass pollen
CC allergen Lol pIa, or an antigenic fragment. The allergen can
CC alternatively be Lol pIb. The antigenic fragment has T-cell
CC stimulating activity and IgE stimulating activity. It does not bind
CC IgE specific for rye grass pollen however. It may be encoded by

```
CC   clone 12R (Q23000) or 26.j (Q22246). Several codons in Q23000 do not
CC   translate into the sequence (R2248) which is printed in the
CC   specification. R22248 corresponds to what is printed.
SQ   Sequence   1239 BP;    282 A;    434 C;    305 G;    217 T;


alignment_scores:
            Quality:  174.00              Length:    325
              Ratio:  1.243                Gaps:     18
 Percent Similarity:  43.077   Percent Identity: 25.231

alignment_block:
 US-09-304-121-2 x Q23000   ..

 Align seg 1/1  to: Q23000  from: 1  to: 1239

     195 SerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleProLe 211
          |||||||||||||:::|||           :::::::     :::|||::
      10 TCCCTCGTACAAACAAAC............GCAAGAGCAGCAATGCCCGT 47

     211 uMetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSerArgG 228
          :                :::   ::::::::::|||   :::   |||:
      48 C.................CAGAAGTACACGGTCCCTCTATTCCTCCGCC 79

     228 lnPheSerLeuGluHisValHisGlySerGlyProTyrSerAlaProSer 244
          ::                          ||||||   ::::::|||:::
      80 GT........................GGCCCTCGTGGCGGGCCCGGC 102

     245 ProAlaTyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSerSe 261
             :::|||::::::   :::   |||:::|||   |||:::|||:::
     103 CGCTCCTACGCCGCTCACGCCGGCTACACCCCCGCAGCCGCGGCCACC.. 150

     261 rGlyProIleIleSerAspIleThrGluLeuAlaProAlaSerProMetA 278
                                          ||||||:::|||   |
     151 .....................................CCGGCTACTCCTGCTG 166

     278 laSerProGlyGlySerIle.........AspGluArg............ 287
          ||:::|||:::|||:::            |||:::|||
     167 CCACCCCGGCTGGCGGCTGGAGGGAAGGCGACGACCGACGAGCAGAAGCT 216

     287 ..................................................... 287

     217 GCTGGAGGACGTCAACGCCTGGCTTCAAGGCAGCCGTGGCCGCCGCTGCC 266

     288 ....ProLeu..................SerSerSerPro.LeuValArg 296
              ||||||                ||||||   ||| |||:::|||
     267 AACGCCCCTCCGGCGGACAAGTTCAAGATCTTCGAGGCCGCCTTCTCCGA 316

     297 ValLysGluGluProProSerProProGlnSerProArgValGluGluAl 313
          |||   :::   ||||||   ||||||      ||||||         ::
     317 GTCCTCCAAGGGCCTCCTCGCCCACCTCCGCCGCCAAGGCACCCGGCCTC 366

     313 aSerProGlyArgProSerSerValAspThrLeuLeuSerProThrAlaL 330
          :|||         |||||||||:::           |||||||||
     367 ATCC.........CCAAGCTCGACACCGCCTACGACGTCGCCTACA.... 403

     330 euIleAspSerIleLeuArgGluSerGluProAlaPro............ 342
                        |||      :::|||:::|||
     404 ................AGGCCGCCGAGGCCCACCCCCGAGGCCAAGTAC 436

     343 ...AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGl 358
            |||   |||||||||||||||:::|||         :::   :::
     437 GACGCCTTCGTCACTGCCCTCACCGAAGCCTCCGNGTCATCGCCGGCGCC 486

     358 uGlyArg......ProProSerProProPro.................T 367
          :::|||      |||   ||||||||||||
     487 CTCGAGGTCCACGCCGTCAAGCCCGCCACCGAGGAGGTCCTCGCTTGCTA 536

     367 hr.SerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGl 383
          |||   |||   :::|||   |||:::
     537 AGATCCCCACCGGTGAGCTGCAGATCGATT..................... 566

     383 uLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnT 400
                    |||   :::||||||
     567 .......GACAAGATCGATGCT............................ 581

     400 hrMetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAsp 416
                           :::|||::::::   ::::::|||      :::
     582 .................GCCTTCAAGATCGCAGCCACCGCC...GCCAAC 611

     417 LeuPheSerProSerValThrValProAspMetSerLeuProAspLeuAs 433
                    |||::::::::::   |||:::      :::|||:::
     612 GCCGCCCCACCAACGATAAGTTCACCGTCTTCGAGAGTGCCTTCAACAAG 661

     433 pSerSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProPro. 449
              |||::::::::::         |||   :::   |||      |||
     662 GCCCTCAATGAGTGCAGCGGCGGCGCTATGAGACCTACAAGTTCATCCCC 711

     450 ..........................................ArgProPro 452
                                                  |||||||||
     712 TCCCTCGAGGCCGCGGTCAAGCAGGCCTACGCCGCCACCGTCGCCCGCCG 761

     453 GluAlaGluAsnSerSerPro 459
          :::      :::|||:::|||
     762 CGCCCGACCTCAAGTACGCCG 782

seq_name: N_Geneseq_36:V02998

seq_documentation_block:
ID   V02998 standard; cDNA; 2898 BP.
AC   V02998;
DT   06-JUL-1998  (first entry)
DE   Mouse neural Mena+ cDNA.
KW   Neural Mena+ gene; mammalian Ena; Enabled gene; Evl gene;
KW   cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW   cell growth; cell motility; mouse; ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             140. .2491
FT                   /*tag= a
PN   WO9801755-A1.
PD   15-JAN-1998.
PF   03-JUL-1997; U11669.
PR   05-JUL-1996; US-675815.
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI   Gertler FB, Niebuhr K, Soriano P, Wehland J;
DR   WPI; 98-101197/09.
DR   P-PSDB; W37151.
PT   Detection of modulators of Mena and Ena-VASP-like genes and proteins
PT   - used in control of cytoskeletal dynamic events in normal and
PT   abnormal cell morphology, adhesion, motility, growth and
PT   differentiation
PS   Example 4; Page 56-57; 77pp; English.
CC   This cDNA comprises novel murine neural Mena+ (mammalian Ena) cDNA
CC   that codes for neural Mena+ protein (see W37151).  It was isolated
CC   from a mouse brain cDNA library using murine Mena cDNA (see V02996)
CC   as probe.  Neural Mena+ contains an exon that introduces 244
CC   amino acids between amino acids 238 and 239 of Mena.  Two other
CC   isoforms, neural Mena++ (see W37152) and neural Mena+++ (see
CC   W37153), are also disclosed.  Unlike Mena, neural Mena isoforms
CC   exhibit neural tissue-specific distribution.  Based on the
CC   disclosed Mena and Evl genes (see also V02996-97) and proteins
CC   (see also W37148-49), a variety of methods and compositions are
CC   provided for screening, isolating and characterising endogenous and
CC   exogenous factors, drugs and therapeutic agents useful to evaluate
CC   and/or control cytoskeletal dynamic events involved in normal and
CC   abnormal cell morphology, adhesion, motility, growth and/or
CC   differentiation.  A method of detecting a modulator of Mena
CC   activity/expression is claimed.
SQ   Sequence   2898 BP;    727 A;    901 C;    692 G;    577 T;


alignment_scores:
            Quality:  170.50              Length:    642
              Ratio:  0.643                Gaps:     28
```

```
 Percent Similarity:  41.277   Percent Identity: 20.717

alignment_block:
 US-09-304-121-2 x V02998   ..

 Align seg 1/1  to: V02998  from: 1  to: 2898

       8 GlyAlaAlaGlyProSerAsnValProAlaPheLeu...ThrLysLeuTr 23
         ||||||||||||      :::      ::::::      ::::::
     125 GGCGCCGCCGGCACCATGAGTGAACAGAGTATCTGTCAGGCAAGAGCTGC 174

      23 pThrLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerProS 40
          :::::::|||   |||         ||||||         |||   |||:
     175 TGTGATGGTCTATGAT.........GATGCCAATAAGAAGTGGGTGCCAG 215

      40 erGlyAsnSer..............PheHisValPhe............ 47
         ::|||:::|||                |||::::::
     216 CTGGTGGCTCAACTGGGTTCAGCAGAGTACATATATATCACCATACAGGC 265

      48 ..................................AspGlnGlyGlnPh 52
                                           |||:::
     266 AACAACACATTCAGAGTTGTGGGCAGAAAGATTCAAGACCATCAGGTTGT 315

      52 eAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAlaSerP 69
             :::   ::::::||||||   :::|||:::|||:::   ::::::
     316 GATAAACTGTGCCATTCCTAAAGGGCTGAAGTACAATCAAGCTACACAGA 365

      69 heValArgGln............LeuAsnMetTyrGlyPheArg...... 79
           :::|||                ::::::||||||::::::
     366 CTTTCCACCAATGGAGGGATGCTAGACAGGTGTATGGTCTCAACTTTGGC 415

      80 ................................................Ly 80
                                                         ::
     416 AGCAAAGAGGATGCCAATGTCTTCGCAAGTGCCATGATGCATGCCTTAGA 465

      80 sValValHisIleGluGlnGlyGlyLeuVal.................. 90
         :|||::::::   :::::::::|||   :::
     466 AGTGTTAAATTCACAGGAAGCAGGGCCAACATTGCCTAGACAAAATTCAC 515

      91 ......................LysProGluArgAspAspThrGluPhe 99
                                |||:::::::::::::   |||:::
     516 AGCTACCTGCTCAAGTTCAAAATGGCCCATCCCAAGAAGAGCTGGAAATC 565

     100 GlnHisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLy 116
         |||:::      :::   :::   |||:::|||      ::::::::::
     566 CAGAGAAGGCAACTGCAAGAACAGCAGCGACAG......AAGGAACTGGA 609

     116 sArgLysValThrSerValSerThrLeuLysSerGluAspIleLysIleA 133
         :|||:::      :::   :::   |||:::   |||   ::::::   |
     610 GAGGGAAAGAATGGAGAGGGAAAGGTTGGAGAGAGAACGACTAGAA...C 656

     133 rgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGly 149
         ||:::          :::   |||   :::   ||||||   :::
     657 GAGAGAGGCTAGAGAGGGAGCGCCTGGAACAAGAGCAGCTGGAGCGGCAG 706

     150 LysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGl 166
         :::|||||||                          :::::::::   ||
     707 CGGCAGGAA.........................AGGGAGCACGTGGA 729

     166 uAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnG 183
         |   |||   ||||||      :::   |||   ::::::   :::|||:
     730 GCGGCTGGAGAGGGAGAGGCTGGAG...CGCCTGGAGCGAGAGAGGCAGG 776

     183 lnLysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSer 199
         :::::      :::::::|||   |||:::      :::   |||:::
     777 AGCGGGAGCGAGAGCGCCTGGAGCAGCTGGAGCGGGAGCAAGTGGAGTGG 826

     200 AsnArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSe 216
         :::|||         :::::::::      |||         :::::::||
     827 GAGCGAGAGCGCAGAATGTCCAATGCTGCTCCA...TCTTCAGACAGCTC 873

     216 rGlySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluH 233
         |   |||:::      :::|||:::||||||
```

```
     874 CCTGTCTAGTGCTCCACTTCCTGAGTATTCC..................... 904

     233 isValHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSer 249
                :::      |||   |||||||||   |||:::|||  :::
     905 ........AGTTGCCAGCCGCCTTCGGCACCTCCTCCATCATAT...GCT 943

     250 SerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSe 266
         :::   :::   ||||||      |||::::::::::   |||      ::
     944 AAAGTCATCTCAGCTCCG......GTGTCAGACGCCACTCCTGATTACGC 987

     266 rAspIleThrGluLeuAlaProAlaSerProMetAlaSerPro....... 280
         :   :::|||   |||   |||:::|||          :::|||
     988 TGTAGTGACTGCTTTGCCACCTACTTCCACACCCCCTACACCACCACTGA 1037

     281 ................................GlyGlySerIleAsp... 285
                                        |||::::::::::
    1038 GACACGCAGCGACACGTTTTGCAACATCTCTAGGTTCAGCCTTCCACCCT 1087

     286 .......................GluArgProLeuSer............ 290
                                 |||||||||:::
    1088 GTTCTTCCCCATTACGCTACAGTTCCTCGTCCTCTCAACAAAAACTCTCG 1137

     291 ....SerSerProLeuValArgValLysGluGluProProSer....... 303
             ||||||||||:::          ::::::::::||||||:::
    1138 ACCTTCTTCTCCTGTGAACACACCCTCTTCTCAGCCTCCAGCTGCGAAGT 1187

     304 ................................ProProGlnSerPro 308
                                          ||||||:::   |||
    1188 CCTGTGCCTGGCCTACTTCCAATTTCTCGCCCCTCCCTCCATCTCCTCCA 1237

     309 ArgValGluGluAlaSerProGlyArgProSerSerValAspThrLeuLe 325
         :::   ::::::   ||||||:::   ::::::         :::||
    1238 ATAATGATTAGCAGCCCCCCTGGCAAAGCTACTGGNCCACGGCCTGTCCT 1287

     325 u.....................................SerProThrA 329
         |                                   ||||||||||
    1288 TCCCGTTTGTGTCTCCTCTCCTGTGCCCCAAATGCCTCCGTCACCAACAG 1337

     329 la.........LeuIleAspSerIleLeuArgGluSerGluProAlaPro 342
         ||          :::||||||:::           :::|||   |||
    1338 CACCCAATGGGTCGCTAGACTCTGTAACATACCCAGTGTCTCCACCGCCT 1387

     343 AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGl 359
         :::|||      |||
    1388 ACCTCAGGGCCAGCA..................................... 1402

     359 yArgProProSerProProProThrSerThrProGluLysCysLeuSerV 376
             ||||||   |||||||||          |||
    1403 .GCGCCACCTCCGCCGCCACCGCCACCGCCGCCG.................. 1435

     376 alAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAsp 392

    1435 ..................................................... 1435

     393 SerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheServa 409

    1435 ..................................................... 1435

     409 lAspThrSerAlaLeuLeuAspLeuPheSerProSerValThrValProA 426
                                          |||         :::|||
    1436 ..........................CCACCACCACCGCCGCTGCCAC 1457

     426 spMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeu 442
             ||||||   |||  :::||||||:::         :::
    1458 CGCCGCCGCTGCCTCCCCTC...GCCTCACTCTCACACTGTGGATCACAG 1504

     443 LeuSerProGlnGluProProArgProProGluAlaGluAsnSerSerPr 459
         ||||||      ||||||      |||   |||      |||:::||
    1505 GCTTCTCCT......CCTCCAGGCACCCCTCTTGCC......TCAACTCC 1542

     459 oAspSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuL 476
         |:::|||                        :::|||      |||
    1543 CTCATCC.........................AAGCCCAGTGTTCTCC 1565
```

```
     476 euAspProGlySerValAspThrGlySerAsnAspLeuProValLeuPhe 492
           :::|||:::::         :::|||   :::   |||:::
    1566 CTTCTCCCTCTGCAGGTGCCCCTGCCTCTGCGGAGACCCCTCTAAATCCT 1615

     493 GluLeuGlyGluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAs 509
         |||||||||::::::|||       :::    |||         |||:::::
    1616 GAGCTGGGAGACTCCTCTGCTTCCGAGCCAGGCTTGCAGGCAGCCTCTCA 1665

     509 pProThrIleSer..............LeuLeuThrGlySerGluProP 521
          :|||:::   |||              |||:::   |||         |
    1666 GCCGGCCGAGTCGCCAACCCCACAGGGCCTTGTCTTGGGACCACCTGCAC 1715

     521 roLysAlaLysAspProThrValSer 529
         ||          |||      |||
    1716 CTCCGCCACCACCCCTCTCCCATCA 1741

seq_name: N_Geneseq_36:T94101

seq_documentation_block:
ID    T94101 standard; DNA; 53526 BP.
AC    T94101;
DT    01-JUN-1998  (first entry)
DE    Human PKD1 gene.
KW    Human; polycystic kidney disease 1; PKD1; treatment;
KW    autosomal dominant polycystic kidney disease; APKD; ss.
OS    Homo sapiens.
PN    WO9744457-A1.
PD    27-NOV-1997.
PF    22-MAY-1997; U08799.
PR    03-JUN-1996; US-658136.
PR    24-MAY-1996; US-655360.
PA    (GENZ ) GENZYME CORP.
PI    Burn T, Connors T, Dackowski W, Germino G, Klinger K,
PI    Qian F;
DR    WPI; 98-018511/02.
PT    Human polycystic kidney disease gene, PKD1 - useful to treat and
PT    diagnose human autosomal or adult onset polycystic kidney disease
PS    Claim 2; Pages 90-118; 257pp; English.
CC    The present sequence is the human polycystic kidney disease 1
CC    (PKD1) gene. The PKD1 gene or polypeptide may be used to treat
CC    autosomal dominant polycystic kidney disease (APKD), and identify
CC    carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.
CC    Antibodies (Ab) that distinguish between normal and mutant PKD1
CC    sequences can also be used in diagnostic tests. Anti-PKD1 Ab may
CC    also be used to perform subcellular and histochemical localisation
CC    studies, and to block the function of PKD1. Ab are also useful in
CC    rational drug design studies to identify and test inhibitors of
CC    PKD1. Sense and antisense sequences derived from the PKD1 gene may
CC    used for detection and therapy.
SQ    Sequence   53526 BP;    8486 A;    17665 C;    15768 G;    11607 T;


alignment_scores:
           Quality:  160.00            Length:    292
             Ratio:  1.185               Gaps:     12
 Percent Similarity:  46.233   Percent Identity: 28.425

alignment_block:
 US-09-304-121-2 x T94101   ..

 Align seg 1/1  to: T94101  from: 1  to: 53526

     209 IleProLeuMetLeuAsnAspSerGlySerAlaHisSerMetPro...Ly 224
         :::||||||:::   :::   :::   |||            :::|||
   34361 CTTCCCCTCCTCCCCTCCCCTACCCCTTCCCCTCCTCCCCTCCCCTAGAC 34410

     224 sTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProTyrS 241
          :::                   |||   :::|||:::|||          |
   34411 CTTCCCCTCACCTCCTCCCGCTGAGCCCCTCCACTCGTCCCCCAGCCCCT 34460

     241 er.AlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAspAl 257
         || :::|||||||||:::              ||||||         |||
   34461 CCCTCCCCTAGCCCCTCCCCTCCCCCTTCCTCCCCTCCTCCCCCTCCCCT 34510

     257 aValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaProA 274
          :::                 |||:::                 |||    |||
   34511 CCTC............CCCCTCCCTCTTCCTCCCCCTCCCCTCCTCCCCC 34548

     274 laSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuSer 290
            |||::::::||||||:::           :::         ||||||
   34549 TTCCTCCCCTCTCCTCCCCCTCCCCTCCTGTCCCCCCTCCTCCCCTCCTC 34598

     291 SerSerProLeuValArgValLysGluGluProPro.SerProProGlnS 307
            ||||||:::                   ||||||   ||||||:::|
   34599 CCTCCTCCCCTCCTCCCCCCTCCTCCTCCCCCTCCTCCCTCCTCCCTCCT 34648

     307 erProArgValGluGluAlaSerProGlyArgProSerSerValAspThr 323
         |||||      ::::::      |||:::   |||         :::   :::
   34649 CCCCCTCCTCCTCCTCCCCTCCTCCCTCCTCCCCTCCTCCCCTCCCCTCC 34698

     324 LeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluPr 340
            ||||||               |||   |||         |||:::||
   34699 TCCCCCTCCCCCCTCCCTTCCTCCCCCTCCCCCTC...CCCTCCTCCCC 34745

     340 oAlaProAlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspT 357
         |:::|||
   34746 CTCTCCT........................................ 34752

     357 hrGluGlyArgProProSerProProProThrSerThrProGluLysCys 373
                     ||||||   |||||||||         |||::::::
   34753 ..........CCTCCCATCCCTCCTCCCATCCCTCCTCCCCGTTCCCAT 34791

     374 LeuSerValAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAl 390
                  :::   :::      :::   |||         |||   ::
   34792 TCTCTCCCCTCCCCCTTCCATTTCTCCCTCCTCCCC......CTGCCCTC 34835

     390 aMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyP 407
          :           :::|||                  ::::::||||||      |
   34836 CTCTCCTCCTCACCTCCCCTTCTCCGCTCCTTTCTTCTCCTCCCTCCCTT 34885

     407 heSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerValThr 423
         |||||         :::   :::   |||   |||||||||||||||
   34886 TCTCTCCTCCCTCCCCTTCTCCCCTTCTCCTCTTCTCCCCTTCTCCTCTC 34935

     424 ValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSer.Ile. 439
               ::::::                |||:::||||||:::   |||  :::
   34936 TTTTCATCCTTCCCTTCTTCCCTCCTTTCCTCCTCTTTTCCCTCTTCTCC 34985

     440 .....GlnGluLeuLeuSerProGlnGluProProArgProProGluAla 454
                     ||||||   |||      ||||||   ||||||:::
   34986 CCCCTCCTCCCCTCCTTCCTCCTCCCATTCCCCCTCCTCCCCCCTCCCAT 35035

     455 GluAsnSerSerProAspSerGlyLysGlnLeuValHisTyrThrAlaGl 471
         :::   |||||||||   |||::::::            ||||||   :::::
   35036 TCCCCCTCCTCCCCTCCTTCCTCCTCC.........CATTACCCCTCCTC 35076

     471 nProLeuPheLeuLeuAspProGlyGerValAspThrGlySerAsnAspL 488
         :|||   :::|||      |||:::                |||||||:::
   35077 TCCTCCCCTCCTCCCACCCCCTCT......CCTCCCGGCTCCTCTCCT. 35119

     488 euProValLeuPheGluLeu 494
            |||:::|||:::   |||
   35120 ..CCCCTCCTCATCCCCCTC 35137

seq_name: N_Geneseq_36:T18551

seq_documentation_block:
ID    T18551 standard; DNA; 53577 BP.
AC    T18551;
DT    06-MAY-1997  (first entry)
DE    Human polycystic kidney disease normal PKD1 gene.
KW    Adult onset polycystic kidney disease; APKD; autosomal dominant;
KW    mutant; transversion; transition; deletion; insertion; ds.
OS    Homo sapiens.
FH    Key             Location/Qualifiers
```

```
FT   misc_feature    4379. .5272
FT                   /*tag= a
FT                   /note= "specifically claimed region of intronless
FT                   cDNA identified by exon trapping"
FT   old_sequence    replace(50652. .50653, cg)
FT                   /*tag= b
FT                   /note= "changes Val codon to Leu codon"
FT   old_sequence    replace(50796. .50797, cg)
FT                   /*tag= c
FT                   /note= "replaces Val codon by Leu codon"
FT   old_sequence    insert(51827. .51828, cc)
FT                   /*tag= d
FT                   /note= "insertion, results in frameshift"
PN   WO9612033-A1.
PD   25-APR-1996.
PF   11-OCT-1995; U13357.
PR   12-OCT-1994; US-323443.
PR   31-JAN-1995; US-381520.
PA   (IGIG-) IG LAB INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Burn TC, Connors TD, Dackowski W, Germino G, Klinger KW;
PI   Landes GM, Qian F;
DR   WPI; 96-222017/22.
PT   Isolated human polycystic kidney disease gene and its mutants -
PT   useful for treatment of polycystic kidney disease and screening for
PT   carriers
PS   Claim 1; Fig 1; 65pp; English.
CC   The present sequence is that of the normal human PKD1 gene from
CC   chromosome 16. Mutations in this gene (e.g. transitions,
CC   transversions, deletions and/or insertions) are associated with
CC   adult-onset polycystic kidney disease (APKD). The PKD1 locus is
CC   GC-rich (62.4%). Comparison of this sequence with a previously
CC   reported partial cDNA sequence revealed differences at three
CC   locations (see features table). The most significant difference is
CC   the presence of two additional cytosine residues on the plus-strand
CC   at position 4566 of the previously reported sequence. The insertion
CC   results in a frame-shift in the predicted protein coding sequence,
CC   leading to replacement of 92 C-terminal amino acids with a novel
CC   12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.
CC   There is a region consisting of 17 tandem copies of a perfect 27 bp
CC   repeat and two large CT-rich regions.
SQ   Sequence   53577 BP;    8495 A;    17681 C;    15785 G;    11616 T;


alignment_scores:
           Quality:  160.00            Length:    292
             Ratio:  1.185               Gaps:     12
 Percent Similarity:  46.233   Percent Identity: 28.425

alignment_block:
 US-09-304-121-2 x T18551   ..

 Align seg 1/1  to: T18551  from: 1  to: 53577

     209 IleProLeuMetLeuAsnAspSerGlySerAlaHisSerMetPro...Ly 224
         :::||||||::::   :::   :::   |||         :::|||
   34357 CTTCCCCTCCTCCCCTCCCCTACCCCTTCCCCTCCTCCCCTCCCCTAGAC 34406

     224 sTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProTyrS 241
          :::                  |||   :::|||:::|||         |
   34407 CTTCCCCTCACCTCCTCCCGCTGAGCCCCTCCACTCGTCCCCCAGCCCCT 34456

     241 er.AlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAspAl 257
         || :::||||||||||:::        ||||||        |||
   34457 CCCTCCCCTAGCCCCTCCCCTCCCCCTTCCTCCCCTCCTCCCCCTCCCCT 34506

     257 aValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaProA 274
          :::           |||:::                  |||   |||
   34507 CCTC...........CCCCTCCCTCTTCCTCCCCCTCCCCTCCTCCCCC 34544

     274 laSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuSer 290
           |||::::::||||||:::      :::         ||||||
   34545 TCCCTCCCCTCTCCTCCCCCTCCCCTCCTGTCCCCCCTCCTCCCCTCCTC 34594

     291 SerSerProLeuValArgValLysGluGluProPro.SerProProGlnS 307
           ||||||:::              ||||||    ||||||:::|
   34595 CCTCCTCCCCTCCTCCCCCCTCCTCCTCCCCCTCCTCCCTCCTCCCTCCT 34644

     307 erProArgValGluGluAlaSerProGlyArgProSerSerValAspThr 323
         |||||     ::::::     |||:::   |||     :::   :::
   34645 CCCCCTCCTCCTCCTCCCCTCCTCCCTCCTCCCCTCCTCCCCTCCCCTCC 34694

     324 LeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluPr 340
           ||||||              |||   |||     |||:::||
   34695 TCCCCCTCCCCCCTCCCTTCCTCCCCCTCCCCCCTC...CCCTCCTCCCC 34741

     340 oAlaProAlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspT 357
         |:::|||
   34742 CTCTCCT..................................... 34748

     357 hrGluGlyArgProProSerProProProThrSerThrProGluLysCys 373
                 ||||||  |||||||||        |||::::::
   34749 ..........CCTCCCATCCCTCCTCCCATCCCTCCTCCCCGTTCCCAT 34787

     374 LeuSerValAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAl 390
               :::   :::      :::   |||         |||   ::
   34788 TCTCTCCCCTCCCCCTTCCATTTCTCCCTCCTCCCC......CTGCCCTC 34831

     390 aMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyP 407
         :         :::|||              ::::::||||||      |
   34832 CTCTCCTCCTCACCTCCCCTTCTCCGCTCCTTTCTTCTCCTCCCTCCCTT 34881

     407 heSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerValThr 423
         |||||      :::   :::   |||   ||||||||||||||
   34882 TCTCTCCTCCCTCCCCTTCTCCCCTTCTCCTCTTCTCCCCTTCTCCTCTC 34931

     424 ValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSer.Ile. 439
               ::::::           |||:::||||||:::   ||| :::
   34932 TTTTCATCCTTCCCTTCTTCCCTCCTTTCCTCCTCTTTTCCCTCTTCTCC 34981

     440 .....GlnGluLeuLeuSerProGlnGluProProArgProProGluAla 454
                   ||||||  |||      ||||||  ||||||:::
   34982 CCCCTCCTCCCCTCCTTCCTCCTCCCATTCCCCCTCCTCCCCCCTCCCAT 35031

     455 GluAsnSerSerProAspSerGlyLysGlnLeuValHisTyrThrAlaGl 471
         :::   |||||||||  |||::::::        ||||||   :::::
   35032 TCCCCCTCCTCCCCTCCTTCCTCCTCC.........CATTACCCCTCCTC 35072

     471 nProLeuPheLeuLeuAspProGlySerValAspThrGlySerAsnAspL 488
         :|||   :::|||      |||:::           |||||||:::
   35073 TCCTCCCCTCCTCCCACCCCCCTCT......CCTCCCGGCTCCTCTCCT. 35115

     488 euProValLeuPheGluLeu 494
           |||:::|||:::   |||
   35116 ..CCCCTCCTCATCCCCCTC 35133

seq_name: N_Geneseq_36:T94108

seq_documentation_block:
ID   T94108 standard; DNA; 53577 BP.
AC   T94108;
DT   01-JUN-1998 (first entry)
DE   Human PKD1 locus between chromosomal markers ATPL (ATP6C) and D16S84.
KW   Human; polycystic kidney disease 1; PKD1; treatment;
KW   autosomal dominant polycystic kidney disease; APKD; ss.
OS   Homo sapiens.
PN   WO9744457-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08799.
PR   03-JUN-1996; US-658136.
PR   24-MAY-1996; US-655360.
PA   (GENZ ) GENZYME CORP.
PI   Burn T, Connors T, Dackowski W, Germino G, Klinger K,
PI   Qian F;
DR   WPI; 98-018511/02.
PT   Human polycystic kidney disease gene, PKD1 - useful to treat and
```

```
PT   diagnose human autosomal or adult onset polycystic kidney disease
PS   Example 5; Pages 60-89; 257pp; English.
CC   The present sequence is the human polycystic kidney disease 1
CC   (PKD1) locus between chromosomal markers ATPL (ATP6C) and D16S84.
CC   The PKD1 gene or polypeptide may be used to treat autosomal
CC   dominant polycystic kidney disease (APKD), and identify carriers
CC   of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
CC   (Ab) that distinguish between normal and mutant PKD1 sequences can
CC   also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
CC   perform subcellular and histochemical localisation studies, and to
CC   block the function of PKD1. Ab are also useful in rational drug
CC   design studies to identify and test inhibitors of PKD1. Sense and
CC   antisense sequences derived from the PKD1 gene may used for
CC   detection and therapy.
SQ   Sequence   53577 BP;    8495 A;    17684 C;    15782 G;    11616 T;


alignment_scores:
           Quality:  160.00            Length:    292
             Ratio:  1.185               Gaps:     12
 Percent Similarity:  46.233   Percent Identity: 28.425

alignment_block:
 US-09-304-121-2 x T94108   ..

 Align seg 1/1  to: T94108  from: 1  to: 53577

     209 IleProLeuMetLeuAsnAspSerGlySerAlaHisSerMetPro...Ly 224
         :::||||||:::   :::   :::   |||         :::|||
   34356 CTTCCCCTCCTCCCCTCCCCTACCCCTTCCCCTCCTCCCCTCCCCTAGAC 34405

     224 sTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProTyrS 241
          :::                  |||   :::|||:::|||         |
   34406 CTTCCCCTCACCTCCTCCCGCTGAGCCCCTCCACTCGTCCCCCAGCCCCT 34455

     241 er.AlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAspAl 257
         || :::||||||||||:::         ||||||         |||
   34456 CCCTCCCCTAGCCCCTCCCCTCCCCCTTCCTCCCCTCCTCCCCCTCCCCT 34505

     257 aValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaProA 274
          :::           |||:::                  |||   |||
   34506 CCTC............CCCCTCCCTCTTCCTCCCCCTCCCCTCCTCCCCC 34543

     274 laSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuSer 290
           |||::::::||||||:::      :::         ||||||
   34544 TTCCTCCCCTCTCCTCCCCCTCCCCTCCTGTCCCCCCTCCTCCCCTCCTC 34593

     291 SerSerProLeuValArgValLysGluGluProPro.SerProProGlnS 307
           ||||||:::              ||||||    |||||||:::|
   34594 CCTCCTCCCCTCCTCCCCCCTCCTCCTCCCCCTCCTCCCTCCTCCCTCCT 34643

     307 erProArgValGluGluAlaSerProGlyArgProSerSerValAspThr 323
         |||||      ::::::      |||:::   |||      :::   :::
   34644 CCCCCTCCTCCTCCTCCCCTCCTCCCTCCTCCCCTCCTCCCCTCCCCTCC 34693

     324 LeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluPr 340
           ||||||              |||   |||      |||:::||
   34694 TCCCCCTCCCCCCTCCCTTCCTCCCCCTCCCCCCTC...CCCTCCTCCCC 34740

     340 oAlaProAlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspT 357
         |:::|||
   34741 CTCTCCT..................................... 34747

     357 hrGluGlyArgProProSerProProProThrSerThrProGluLysCys 373
               ||||||   |||||||||         |||:::::
   34748 ...........CCTCCCATCCCTCCTCCCATCCCTCCTCCCCGTTCCCAT 34786

     374 LeuSerValAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAl 390
               :::   :::      :::   |||         |||   ::
   34787 TCTCTCCCCTCCCCCTTCCATTTCTCCCTCCTCCCC......CTGCCCTC 34830

     390 aMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyP 407
         :         :::|||              ::::::||||||         |
   34831 CTCTCCTCCTCACCTCCCCTTCTCCGCTCCTTTCTTCTCCTCCCTCCCTT 34880

     407 heSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerValThr 423
         |||||      :::   :::   |||   ||||||||||||||||
   34881 TCTCTCCTCCCTCCCCTTCTCCCCTTCTCCTCTTCTCCCCTTCTCCTCTC 34930

     424 ValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSer.Ile. 439
               ::::::            |||:::||||||:::   ||| :::
   34931 TTTTCATCCTTCCCTTCTTCCCTCCTTTCCTCCTCTTTTCCCTCTTCTCC 34980

     440 .....GlnGluLeuLeuSerProGlnGluProProArgProProGluAla 454
                 ||||||   |||      ||||||   ||||||:::
   34981 CCCCTCCTCCCCTCCTTCCTCCTCCCATTCCCCCTCCTCCCCCCTCCCAT 35030

     455 GluAsnSerSerProAspSerGlyLysGlnLeuValHisTyrThrAlaGl 471
         :::   ||||||||   |||::::::         ||||||   :::::
   35031 TCCCCCTCCTCCCCTCCTTCCTCCTCC.........CATTACCCCTCCTC 35071

     471 nProLeuPheLeuLeuAspProGlySerValAspThrGlySerAsnAspL 488
         :|||   :::|||      |||:::            |||||||:::
   35072 TCCTCCCCTCCTCCCACCCCCCTCT......CCTCCCGGCTCCTCTCCT. 35114

     488 euProValLeuPheGluLeu 494
           |||:::|||:::   |||
   35115 ..CCCCTCCTCATCCCCCTC 35132

seq_name: N_Geneseq_36:Q53866

seq_documentation_block:
ID   Q53866 standard; cDNA; 4297 BP.
AC   Q53866;
DT   21-JUN-1994  (first entry)
DE   Ge protein-encoding cDNA.
KW   Ge protein; HeLa cell; nuclear; autoantigen; detection;
KW   autoimmune; disorder; disease; Sjorgrens syndrome;
KW   antinuclear antibody; viral infection; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             2. .3457
FT                   /*tag=  a
FT                   /product= Partial_Ge_protein
PN   US5272256-A.
PD   21-DEC-1993.
PF   03-FEB-1992; 844298.
PR   03-FEB-1992; US-844298.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Bloch DB;
DR   WPI; 93-413471/51.
DR   P-PSDB; R43950.
PT   Ge protein derived from HeLa cells - used to detect and predict
PT   auto-immune disorders eg Sjogren's syndrome
PS   Disclosure; Fig 2; 20pp; English.
CC   This sequence encodes a fragment of Ge protein.  The full length Ge
CC   protien has a molecular weight of 170 kD when derived from HeLa
CC   cells.  Ge protein is a nuclear autoantigen that may be used to
CC   detect autoimmune disorders such as Sjorgrens syndrome, which is
CC   characterised by the presence of antinuclear antibodies binding to
CC   the Ge protein.  The protein may be used to detect a patient at risk
CC   from an autoimmune disorder, to prevent autoimmune disease, and to
CC   detect viral infection.
SQ   Sequence   4297 BP;    907 A;    1289 C;    1251 G;    850 T;


alignment_scores:
           Quality:  155.00            Length:    450
             Ratio:  0.735               Gaps:     21
 Percent Similarity:  46.889   Percent Identity: 24.000

alignment_block:
 US-09-304-121-2 x Q53866   ..

 Align seg 1/1  to: Q53866  from: 1  to: 4297
```

```
 105 LeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSe 121
1022 CTCTCAGCAGAATACCTGATTCTCAGCGATGTGCAACGGAAGGTCCTCTA 1071

 121 rValSerThrLeu......LysSerGluAspIleLysIleArgGlnAspS 136
1072 TGTGATGGAGCTGCTGCAAAACCAGGAGGAGGGCCACGCCTGCTTCAGCT 1121

 136 erValThrLys...LeuLeuThrAspValGlnLeuMetLysGlyLysGln 151
1122 CCATCTCGGAGTTCCTGCTCACCCACCCTGTGCTGAGCTTTGGTATCCAG 1171

 152 ..............GluCysMetAspSerLysLeuLeuAlaMetLysHi 163
1172 GTTGTGAGTCGCTGCCGGCTACGGCACACTGAGGTGCTGCCTGCCGAAGA 1221

 163 sGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisA 180
1222 GGAAAATGACAGCCTGGGT.........GCTGATGGTACCCATGGAGCCG 1262

 180 laGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleSerLeu 196
1263 GTGCCATGGAGTCTGCGGCCGGTGTGCTCATCAAG......CTCTTTTGT 1306

 197 ValGlnSerAsnArgIleLeuGlyValLysArgLysIleProLeuMetLe 213
1307 GTGCATACTAAGGCACTGCAAGATGTGCAGATCCGCTTCCAGCCACAGCT 1356

 213 uAsn.....................AspSerGlySerAlaHis....... 220
1357 GAACCCTGATGTGGTGGCCCCACTGGGGACCCACACTGCCCACGAGGACT 1406

 221 ..............SerMetProLysTyrSerArgGlnPheSerLeuGlu 232
1407 TCACATTTGGAGAGTCTCGGCCCGAACTGGGCTCTGAGGGCCTGGGGTCA 1456

 233 HisValHisGlySerGlyPro.....................TyrSerAl 242
1457 GCCGCTCACGGCTCCCAGCCTGACCTCCGACGAATCGTGGAGCTGCCTGC 1506

 242 aProSerProAlaTyrSerSerSerSer..............LeuTyrA 254
1507 ACCTGCCGACTTCCTCAGTCTGAGCAGTGAGACCAAGCCCAAGTTGATGA 1556

 254 laProAspAlaValAlaSerSerGlyProIleIleSerAspIleThrGlu 270
1557 CACCTGACGCCTTCATGACACCTAGCGCCTCCTTGCAGCAGATCACT... 1603

 271 LeuAlaProAlaSerProMetAlaSerProGlyGlySerIleAspGluAr 287
1604 .........GCCTCTCCCAGCAGCAGCAGCAGCGGTAGCAGCAGCAGCAG 1644

 287 gProLeuSerSerSerProLeuValArgVal..................L 298
1645 CAGCAGTAGCAGCAGCTCCCTTACAGCTGTGTCTGCCATGAGCAGCACCT 1694

 298 ysGluGluProProSer.........ProProGln............Ser 307
1695 CAGCTGTGGACCCCTCCTTGACCAGGCCACCTGAGGAGCTGACCTTGAGC 1744

 308 ProArgValGlu...GluAlaSerProGlyArgProSerSerValAspTh 323
1745 CCCAAGCTGCAGCTGGATGGCAGCCTGACAATGAGCAGCAGTGGCAGCCT 1794

 323 rLeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluP 340
1795 TCAGGCAAGCCCGCGTGGCCTCCTGCCTGGCCTGCTC............C 1832

 340 roAlaProAla..................................... 343
1833 CAGCCCCAGCTGACAAACTGACTCCCAAGGGGCCGGGCCAGGTGCCTACT 1882

 344 SerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyAr 360
1883 GCCACCTCTGCACTGTCCCTGGAGCTGCAGGAAGTGGAGCCCCTGGGGCT 1932

 360 gProProSerProProProThrSerThrProGluLysCysLeuSerValA 377
1933 ACCCCAAGCCTCCCCTAGCCGCACTCGTTCCCCTGATGTCATCTCCTCAG 1982

 377 laCysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSer 393
1983 CTTCCACTGAC..............................CTGTCCCAG 2002

 394 AsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAs 410
2003 GACATCCCTGAGATTGCATCTGAGGCCCTGTCCCGTGGTTTT........ 2044

 410 pThrSerAlaLeuLeuAspLeuPheSerProSerValThrValProAspM 427
2045 ............................GGCTCCTCTGCACCAGAGG 2063

 427 etSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLeu 443
2064 GCCTTGAGCCAGACAGTATGGCTTCAGCCGCCTCGGCACTGCACCTGCTG 2113

 444 SerProGlnGluProProArgProProGluAlaGluAsnSerSerProAs 460
2114 TCCCCACGG......CCCCGGCCA..................GGGCCCGA 2139

 460 pSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuA 477
2140 GCTCGGCCCCCAGCTCGGG.........................CTTG 2162

 477 spProGlySerValAspThrGlySerAsnAspLeuProValLeuPheGlu 493
2163 ATGGAGGCCCTGGGGATGGAGATCGGCATAATACCCCCTCCCTCCTGGAG 2212
```

seq_name: N_Geneseq_36:V05287

seq_documentation_block:

```
ID   V05287 standard; DNA; 49377 BP.
AC   V05287;
DT   21-MAY-1998  (first entry)
DE   The soraphen biosynthesis gene cluster from Sorangium cellulosum.
KW   Polyketide synthase; PKS; biosynthesis; soraphen; SorR; SorA; SorB;
KW   SorM; biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW   ketoreductase; beta-ketone processing domain; cytostatic agent;
KW   antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW   biological control; ss.
OS   Sorangium cellulosum.
FH   Key             Location/Qualifiers
FT   CDS             383. .760
FT                   /*tag= a
FT                   /product= SorR
FT                   /note= "gene product highly homologous to the
FT                         reductase domains of type I PKSs such
FT                         as eryA from Saccharopolyspora erythraea"
FT   CDS             927. .19874
FT                   /*tag= b
FT                   /product= SorA
FT                   /note= "gene product is highly homologous to
FT                         type I PKSs that are known to be involved
FT                         in the synthesis of polyketide compounds"
FT   misc_feature    942. .7115
FT                   /*tag= c
FT                   /note= "module 1 of SorA"
FT   misc_feature    7203. .12884
FT                   /*tag= d
FT                   /note= "module 2 of SorA"
FT   misc_feature    13455. .19616
FT                   /*tag= e
FT                   /note= "module 3 of SorA"
FT   misc_feature    19871. .46318
FT                   /*tag= f
FT   CDS             19871. .46318
```

```
FT                   /*tag=  g
FT                   /product= SorB
FT                   /note= "gene product is highly homologous to
FT                          type I PKS genes"
FT   misc_feature    19870. .24556
FT                   /*tag=  h
FT                   /note= "module 1 of SorB"
FT   misc_feature    24638. .30820
FT                   /*tag=  i
FT                   /note= "module 2 of SorB"
FT   misc_feature    30881. .35446
FT                   /*tag=  j
FT                   /note= "module 3 of SorB"
FT   misc_feature    35528. .40114
FT                   /*tag=  k
FT                   /note= "module 4 of SorB"
FT   misc_feature    40190. .46318
FT                   /*tag=  l
FT                   /note= "module 5 of SorB"
FT   CDS             46851. .47891
FT                   /*tag=  m
FT                   /product= SorM
FT                   /note= "gene product is homologous to the
FT                          methyltransferase from Streptomyces
FT                          hygroscopicus that is involved in
FT                          the synthesis of the polyketide rappamicin"
PN   US5716849-A.
PD   10-FEB-1998.
PF   14-DEC-1996; 764233.
PR   24-AUG-1993; WO-U07954.
PR   08-JUN-1994; US-258261.
PR   09-OCT-1996; US-729214.
PA   (NOVS ) NOVARTIS FINANCE CORP.
PI   Beck JJ, Hill DS, Ligon JM, Neff S, Ryals JA, Schupp T;
DR   WPI; 98-158369/14.
PT   DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT   biosynthesis of soraphen useful as antimicrobial agent against
PT   phytopathogenic fungi
PS   Claim 2; Columns 47-90; 64pp; English.
CC   The present sequence contains a cluster of genes that encode polyketide
CC   synthases (PKSs) that are involved in the synthesis of soraphens in
CC   Sorangium cellulosum. The proteins encoded by the present sequence are
CC   SorR, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC   which contain a beta-ketoacylsynthase, an acyltransferase, a
CC   ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC   processing domains. S. cellulosum soraphens are useful as a cytostatic
CC   and antimicrobial agent active against phytopathogenic fungi.
CC   Soraphen-producing transgenic plants or biological control agents can
CC   also be produced, which may reduce crop losses and nutritional
CC   deprivation for local populations in many parts of the world.
SQ   Sequence   49377 BP;    7247 A;    19522 C;    14477 G;    8131 T;


alignment_scores:
            Quality:  155.00              Length:    416
              Ratio:  0.799                Gaps:     18
 Percent Similarity:  46.635   Percent Identity: 26.442

alignment_block:
 US-09-304-121-2 x V05287   ..

 Align seg 1/1  to: V05287  from: 1  to: 49377

     139 LysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAs 155
         |||   |||   |||      |||   |||::::::
    8598 AAGCGGCTCCGCGACCACCTCCTCGCCAAAAGCGAG..............  8633

     155 pSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluV 172
                 ||||||:::               :::|||
    8634 .........CTCGCCCTCGCCGACGTGGCCTATTCGCTCGCGACCACGC 8673

     172 alAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLys 188
           |||   ::::::|||:::   |||         |||      ::::::

    8674 GCGCCCACTTCGAGCAGCGCGCCGCTCTCCTCGTCAAAGGCCGCGACGAG 8723

     189 LeuIleGlnPheLeuIleSerLeuValGln......SerAsnArgIleLe 203
         |||::::::   |||   ||||||:::|||      |||      :::||
    8724 CTCCTCTCCGCCCTCGATTCGCTGGCCCAAGGACATTCCGCCGCCGTGCT 8773

     203 u.........GlyValLysArgLysIleProLeuMetLeuAsnAspSerG 217
         |         |||:::      |||:::   ::::::::::::   :::|
    8774 CGGACGAAGCGGCGCCCCAGGAAAGCTCGCCGTCCTCTTCACGGGGCAAG 8823

     217 lySerAlaHis...SerMetProLysTyr.SerArgGlnPheSerLeuGl 232
         |||||   :::   :::|||   :::    |||      :::|||
    8824 GAAGCCAGCGGCCCACCATGGGCCGCGCCCTCTACGACGCTTTCC..... 8868

     232 uHisValHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerS 249
                 :::||||||      |||:::|||||||||      |||:
    8869 .......CCGTCTTCCGGGACGCCCTCGACACCGTCGCCGCCCACCTCGA 8911

     249 erSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIle 265
         ::::::::      ||||||   |||:::::::||||||   |||
    8912 CCGCGACCTCGACCGCCCCCTGCGCGACGTCCTCTTCGCTCCCG...... 8955

     266 SerAspIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGl 282
                              ||||||:::   |||   ||||||
    8956 ..............ACGGCTCCGAGCAGGCCGCGCGCCTCG........ 8982

     282 ySerIleAspGluArgProLeuSerSerSerProLeuValArgValLysG 299

    8983 .............................................A 8983

     299 luGluProProSerProProGlnSerProArgValGluGluAlaSerPro 315
           :::||||||||||||||   :::   |||   :::   :::   ||||||
    8984 CCAAACCGCCTTCACCCAGCCGG...CCCTGTTTGCCCTCGAAGTCGCCC 9030

     316 GlyArgProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAs 332
         :::         |||       ::::::   ||||||   :::      ||
    9031 TCTTTCAGCTTCTCCAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGA 9080

     332 pSerIleLeuArgGluSerGluPro......AlaProAlaSerValThrA 347
         |:::   |||   :::|||:::|||      :::||||||||      :
    9081 CACTCCATTGGCGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCT 9130

     347 laLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProProSer 363
         ::   |||   |||                              |||   |||
    9131 CCAGGACGCCTGCA.........................CCCTCGTCG 9153

     364 ProProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAs 380
         ||||||      |||:::   ::::::      ::::::|||      ::
    9154 CCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTC 9203

     380 pLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn......L 395
         :   :::|||         :::   :::|||      |||:::      :
    9204 ACCCTCCGAGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGA 9253

     395 euAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThr 411
         ::|||:::|||   :::         |||   |||:::|||      :::
    9254 TGGACGAGCTAGCCTCGCCG...CCCTCAATGGGCCTCTCTCCACCGTCG 9300

     412 SerAlaLeuLeuAspLeu.........PheSerProSerValThrValPr 425
         |||      :::                ||||||:::             ||
    9301 TCGCTGGCGATGAAGACGCGGTGGTGGAGATCGCCCGCCAGGCCGAAGCC 9350

     425 oAspMetSerLeuProAspLeuAspSerSerLeu................ 436
         |:::   :::   |||          :::|||
    9351 CTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTTCCACTCTCC 9400

     437 .............AlaSerIleGlnGluLeuLeuSerProGlnGluPro 448
                      :::|||      :::      ||||||:::
    9401 GCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 9450

     449 ProArgProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLe 465
         |||      |||:::|||:::   |||||||||   |||      :::
    9451 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGC 9500
```

```
 465 uValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValA 482
        :::   |||::::::|||                   ::::::
9501 GCCACGGACCACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCG 9550

 482 spThr......GlySerAsnAspLeuProValLeuPhe...GluLeuGly 495
       |||        :::|||:::         |||   |||       |||
9551 CCACACCGTCCGCTTCCTCGACGGCGTACGTGCCCTTCACGCCGAAGGGG 9600

 496 GluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAspProThrIl 512
     ::::::|||:::   |||:::|||
9601 CACGCGTCTTTCTCGAGCTCGGGC.......................... 9624

 512 eSerLeuLeuThrGlySerGluProProLysAlaLysAspProThr 527
            ||||||   |||:::||||||       |||   |||:::
9625 .......CTCACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCG 9663
```

seq_name: N_Geneseq_36:N90357

```
seq_documentation_block:
ID   N90357 standard; DNA; 2589 BP.
AC   N90357;
DT   02-NOV-1992  (first entry)
DE   Sequence of a genetic construct which encodes CD4 linked to human
DE   IgG1 at the Esp site upstream of the hinge region (fusion protein
DE   CD4E-gamma-1).
KW   Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW   diagnosis; CD4; gp120; binding fragment; glycoprotein;
KW   variable region; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon            111. .1307
FT                   /*tag= a
FT   intron          1308. .1664
FT                   /*tag= b
FT                   /note= "1"
FT   exon            1665. .1709
FT                   /*tag= c
FT   intron          1710. .1827
FT                   /*tag= d
FT   exon            1828. .2157
FT                   /*tag= e
FT   intron          2158. .2253
FT                   /*tag= f
FT   exon            2254. .2577
FT                   /*tag= g
PN   EP-325262-A.
PD   26-JUL-1989.
PF   20-JAN-1989; 100913.
PR   22-JAN-1988; US-147351.
PA   (GEHO-) GEN HOSPITAL CORP.
PI   Seed B;
DR   WPI; 89-214472/30.
DR   P-PSDB; P93009.
PT   Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT   infections or detecting HIV or SIV in sample
PS   Example; Table 2, Page 24-33; 68pp; English.
CC   The fusion protein genes of the invention pref. comprise cDNA
CC   sequences which encode CD4 or a fragment which binds gp120 ligated
CC   to an expression plasmid which encodes an antibody in which the
CC   variable region of the gene has been deleted (see WO87-02671).  The
CC   CD4 portion of the fusion protein may comprise the complete CD4
CC   sequence, the 370 AA extracellular region and the membrane spanning
CC   domain, or the extracellular region.  The Ig heavy chain is pref.
CC   from IgM, IgG1 or IgG3.  The following are specifically claimed:
CC   fusion proteins CD4H-gamma-1,CD4Mmu,CD4Pmu,CD4E-gamma1,and CD4Mmu
CC   (No.67608),pCD4P-gamma (No.67609) and pCD4E-gamma-1 (No.67610).
CC   The plasmid containing (pCD4E-gamma-1) has been deposited in E.
CC   coli (MC1061/P3) at the ATCC under accession number 67610.
SQ   Sequence   2589 BP;   618 A;   793 C;   712 G;   465 T;
```

alignment_scores:

```
            Quality:  148.00             Length:    444
              Ratio:  0.722                Gaps:     22
 Percent Similarity:  46.171   Percent Identity: 22.973
```

```
alignment_block:
 US-09-304-121-2 x N90357   ..

 Align seg 1/1  to: N90357  from: 1  to: 2589

  24 ThrLeuValSerAspProAspThrAspAlaLeuIle...CysTrpSerPr 39
     ||||||   |||   |||   ::::::        :::   |||   |||||
 534 ACCTTGGAGAGCCCCCCTGGTAGTAGCCCCTCAGTGCAATGTAGGAGTCC 583

  39 oSerGly..............................AsnSerPheH 45
     |    |||                                ::::::::::
 584 AAGGGGTAAAAACATACAGGGGGGGAAGACCCTCTCCGTGTCTCAGCTGG 633

  45 isValPheAspGlnGlyGlnPheAlaLysGluValLeuProLysTyrPhe 61
     :::::    |||:::|||    ::::::       ||||||
 634 AGCTCCAGGATAGTGGCACCTGGACATGCACTGTCTTG............ 671

  62 LysHisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPh 78
     ::::::::::::    ::::::|||        ::::::   :::    :::||
 672 CAGAACCAGAAGAAGGTGGAGTTCAAAATAGACATCGTGGTGCTAGCTTT 721

  78 eArgLysValValHisIleGluGlnGlyGlyLeuValLysProGluArgA 95
     |:::|||:::        |||             :::   |||   |||    :
 722 CCAGAAGGCCTCCAGCATA...........GTCTATAAGAAAGAGGGGG 759

  95 spAspThrGluPheGlnHisProCysPhe..............LeuArg 106
     ::::::::||||||:::    |||                     |||
 760 AACAGGTGGAGTTCTCCTTCCCACTCGCCTTTACAGTTGAAAAGCTGACG 809

 107 GlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerValSe 123
     |||:::    :::|||        :::    :::|||        :::|||    ||
 810 GGCAGTGGCGAGCTGTGGTGGCAGGCGGAGAGGGCTTCCTCCTCCAAGTC 859

 123 rThrLeuLysSerGluAspIleLysIleArgGlnAspSerValThrLysL 140
     |    :::    :::    |||:::|||     ::::::    ||||||     ::::
 860 TTGGATC...ACCTTTGACCTGAAGAACAAGGAAGTGTCTGTAAAACGGG 906

 140 euLeuThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAspSer 156
     ::         |||    :::|||:::    |||||||:::          :::
 907 TTACCCAGGACCCTAAGCTCCAGATGGGCAAGAAGCTCCCGCTCCACCTC 956

 157 LysLeu.......................................Le 159
        |||                                         ||
 957 ACCCTGCCCCAGGCCTTGCCTCAGTATGCTGGCTCTGGAAACCTCACCCT 1006

 159 uAlaMetLysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuA 176
     ||||::::::    ::::::        |||    :::||||||       :::
1007 GGCCCTTGAAGCGAAAACAGGAAAGTTGCATCAGGAAGTGAACCTGGTGG 1056

 176 rgGlnLysHisAlaGlnGlnGlnLysValValAsnLysLeuIle...... 190
       ::::::    :::|||    ||||||    ::::::      :::
1057 TGATGAGAGCCACTCAGCTCCAGAAAAATTTGACCTGTGAGGTGTGGGGA 1106

 191 ...........GlnPheLeuIleSerLeuValGlnSerAsnArgIleLe 203
                ::::::::::::||||||      :::|||:::
1107 CCCACCTCCCCTAAGCTGATGCTGAGCTTGAAACTGGAGAACAAGGAGGC 1156

 203 uGlyVal...LysArgLysIleProLeu..................... 211
         |||    ||||||:::   |||:::
1157 AAAGGTCTCGAAGCGGGAGAAGCCGGTGTGGGTGCTGAACCCTGAGGCGG 1206

 212 .............MetLeuAsnAspSerGly................. 217
                   :::|||:::|||||||||
1207 GGATGTGGCAGTGTCTGCTGAGTGACTCGGGACAGGTCCTGCTGGAATCC 1256

 218 SerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisVa 234
     :::          :::|||    :::|||                  |||::
1257 AACATCAAGGTTCTGCCCACATGGTCCACCCCGGTG.........CACGC 1297
```

```
     234 lHisGlySerGlyProTyrSerAlaProSer.................P 245
         :      :::|||   |||   |||   :::                  |
    1298 GGATCCCGAGGGTGAGTACTAAGCTTCAGCGCTCCTGCCTGGACGCATCC 1347

     245 roAlaTyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSerSer 261
         |||||  :::  :::|||      |||  :::|||  ::::::|||
    1348 CGGCTATGCAGCCCCAGTCCAGGGCAGCAAGGCAGGCCCCGTCTGCCTCT 1397

     262 GlyPro.....................................IleIl 265
         :::|||                                     :::::
    1398 TCACCCGGAGCCTCTGCCCGCCCCACTCATGCTCAGGGAGAGGGTCTTCT 1447

     265 eSerAspIleThrGluLeu.................AlaProAlaSerP 276
         ::::           |||                 ||||||
    1448 GGCTTTTTCCCAGGCTCTGGGCAGGCACAGGCTAGGTGCCCCTAACCCAG 1497

     276 roMetAlaSerProGlyGlySerIleAspGluArgProLeuSerSerSer 292
           |||     |||::::::        ||||||  :::|||
    1498 GCCCTGCACACAAAGGGGCAGGTGCTGGGCTCAGACCTGCCAAGAGC... 1544

     293 ProLeuValArgValLysGluGluProProSerProProGlnSerProAr 309
             :::::::::|||:::  |||:::|||     ||||||
    1545 .........CATATCCGGGAGGAC...CCTGCCCCTGACCTAAGCCCA.. 1580

     309 gValGluGluAlaSerProGlyArgProSerSer.............ValA 322
                   |||  ||||||:::|||            :::
    1581 ...............CCCCAAAGGCCAAACTCTCCACTCCCTCAGCTCG 1614

     322 spThrLeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSer 338
           |||||||||           :::        |||  ::::::
    1615 GACACCTTCTCTCCTCCCAGATTCCAGTAACTCCCAATCTTCTCTCTGCA 1664

     339 GluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGlyHisTh 355
         ||||||     ::::::  |||                  |||||
    1665 GAGCCCAAATCTTGTGACAAAACT.....................CACAC 1693

     355 r......AspThrGluGlyArgProProSerProProProThrSerThr. 369
         |              |||:::|||  :::     |||:::|||:::
    1694 ATGCCCACNGTGCCCAGGTAAGCCAGCCCAGGCCTCGCCCTCCAGCTCAA 1743

     370 ..ProGluLysCysLeuSerValAlaCysLeu 379
           ::::::|||     ||||||||||:::
    1744 GGCGGGACAGGTGCCCTAGAGTAGCCTGCATC 1775

seq_name: N_Geneseq_36:T06769

seq_documentation_block:
ID   T06769 standard; DNA; 28598 BP.
AC   T06769;
DT   15-OCT-1996  (first entry)
DE   Sorangium cellulosum soraphen gene cluster.
KW   Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;
KW   fungicide; pesticide; myxobacterium; ss.
OS   Sorangium cellulosum.
PN   WO9533818-A2.
PD   14-DEC-1995.
PF   30-MAY-1995; IB0414.
PR   08-JUN-1994; US-258261.
PA   (CIBA ) CIBA GEIGY AG.
PI   Beck JJ,  Gaffney TD,  Hammer PE,  Hill DS,  Lam ST;
PI   Ligon J,  Ryals JA, Schupp T,  Uknes SJ;
DR   WPI; 96-040226/04.
PT   New genes for biosynthesis of anti-pathogenic substances - pref.
PT   pyrrolnitrin and soraphen, useful for disease control in plants
PS   Claim 3; Page 140-158; 190pp; English.
CC   This is the soraphen gene cluster from S. cellulosum.  The gene
CC   cluster may be expressed recombinantly to produce soraphen, or
CC   expressed in a transgenic plant for disease-resistance.
SQ   Sequence   28598 BP;    4031 A;    11703 C;    8263 G;    4601 T;


alignment_scores:
           Quality:  146.50            Length:    527
                Ratio:  0.654                Gaps:     29
 Percent Similarity:  42.505   Percent Identity: 22.201

alignment_block:
 US-09-304-121-2 x T06769   ..

 Align seg 1/1  to: T06769  from: 1  to: 28598

     155 AspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGl 171
         ::::::  ::::::   :::  ::::::  |||||||||  |||::
    5761 GAAGACGCGGTGGTGGAGATCGCCCGCCAGGCCGAAGCCCTCGGACGAAA 5810

     171 uValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnL 188
         :::::::  ||||||  :::||||||  ::::::      ::::::
    5811 GACCACACGCCTGCGCGTCAGCCACGCCTTCCATTCCCCGCACATGGACG 5860

     188 ysLeuIleGlnPheLeuIleSerLeuValGlnSerAsnArgIleLeuGly 204
           :::::::::  :::      ::::::||||||     :::
    5861 GAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGC.....CTCACCTAC 5904

     205 ValLysArgLysIleProLeuMetLeuAsnAspSerGlySerAla..... 219
               :::||||||:::::  |||  :::|||:::
    5905 CATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCAC 5954

     220 ....His.SerMetProLysTyrSerArgGlnPheSerLeuGluHisVal 234
             ||| |||  |||:::  :::
    5955 GGACCACGAGCTCGCCTCGCCCGACTACT..................... 5983

     235 HisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSerSe 251
           ||||||:::  :::::::::||||||        :::||||||::
    5984 ...GGGTCCGCCACGTTCGCCACACCGTCC.........GCTTCCTCGAC 6021

     251 rLeuTyrAlaPro.................AspAlaValAlaSerSerG 262
         :  |||:::|||                :::  :::|||||||
    6022 GGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGG 6071

     262 lyProIleIle.........................SerAspIleThr 269
         ||      :::                       ||||||  |||
    6072 GCCTCACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACG 6121

     270 GluLeu.........AlaProAla........................... 274
         :::            ||||||:::
    6122 AAGGCACGTCGCCATGCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGAC 6171

     275 .SerProMetAlaSerProGlyGlySerIleAspGluArgProLeuSerS 291
          :::|||      ||||||     |||      :::  |||  :::|
    6172 GACGCCGAGGCGTTCACCGCCGCGCTCGGCGCTCTCCACTCCGCAGGCAT 6221

     291 erSerProLeuValArgValLysGluGluProProSerProProGlnSer 307
         ||  |||  ::::  :::  :::::|||||||||||||||||:::
    6222 CACACCCG...ACTGGAGCGCTTTCTTCGCCCCCTTCGCCTCCAAGGCAC 6268

     308 ProArgValGluGluAlaSer.....................ProGlyAr 317
         |||          :::|||:::                      |||::
    6269 CCGGCGCCGACGTCAGCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAA 6318

     317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
         :|||||||||:::  |||      |||      :::  :::::::|||
    6319 GCCATCGAGCGCGGGGACC......TCGATGCGCTCAGCGGTCAGCTCCA 6362

     334 leLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAsp 350
                 ::::::  ||||||  |||      ::::::
    6363 CGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCCTTCCCACCC 6412

     351 AlaArg.....................................Gl 353
         :::|||                                     ||
    6413 TCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCAGCACGGTCGACGCCTGG 6462

     353 yHisThrAspThrGluGlyArg.........ProProSerProProProT 367
         |  |||  :::  |||           ||||||:::  |||||||
    6463 CGCTACCGTATCACCTGGAAGCCTCTGACCACCGCCGAAACACCCGCCGA 6512

     367 hrSer..................................... 368
```

```
     |||||
6513 CCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACGACG 6562

 369 ..............ThrProGluLys..................... 372
                   :::|||::::::
6563 CGCTCCCCTCCGCGCTCACCGAGGCGCTCACCCGGCGCGGCGCGCGCGTC 6612

 373 .......Cys..LeuSerValAlaCysLeuAspLysAsnGluLeuSerAs 386
          |||  ||||||   |||   ||||||::::::   |||:::::
6613 CTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGGCTCTCGCCGA 6662

 386 pHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuS 403
     :||||||           ::::::         ::::::   :::||||
6663 GCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGATTCGCGGCGTGCTCT 6712

 403 er.SerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSe 419
     || |||                                        ||
6713 CGCTCC.....................................TC 6720

 419 rProSerValThrValProAspMetSerLeu.....ProAspL 432
     |||||||:::::::::|||::::::::::::      ||||||
6721 GCCCTCGACGAGCGCCCCCTCGCAGACCGTCCTGCCCTGCCCGCCGGACT 6770

 432 euAspSerSerLeuAlaSerIleGln..................... 440
          |||   :::   |||::::::
6771 CGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCGACCTCGAGG 6820

 441 ..................................GluLeuLeuSe 444
                                      :::   |||::
6821 CGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGAC 6870

 444 rProGlnGlu...ProProArgPro.............ProGluA 454
     :|||:::      ||||||||||||                ::::
6871 CCCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCAT 6920

 454 laGluAsnSerSerProAspSerGlyLysGlnLeuValHisTyrThrAla 470
     ::   :::|||:::|||   :::|||:::      |||      |||:::
6921 CGGCCTCGAGCACCCCGACCGGTGGGGAG......GTCTCGTCGACGTCT 6964

 470 ................................................ 470

6965 GCGCTGGGGTCGACGAGAGCGCCGTGGGCCGCTTGCTGCCGGCCCTCGCC 7014

 471 ................GlnProLeuPheLeuLeuAspPro...GlyS 480
                     ||||||   |||   :::|||   ||||
7015 GAGCGCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCGCCCGATGGCT 7064

 480 erValAspThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu 496
     ||:::                                    |||
7065 CGCTC.....................................GAA 7072

 497 GlySerTyrPheSerGluGlyAspGlyPheAlaGluAspPro........ 510
        :::      ::::::::::::::   ||||||   |||
7073 GAGGCGCTCAGCACCTCGTCCTCATCAGCCGCCGAGGCGCCGAGGCCCCT 7122

 511 ....ThrIleSerLeuLeuThrGlySerGluPro............... 520
             |||      |||:::|||:::|||
7123 GGCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGCGCACCAC 7172

 521 .....ProLysAlaLysAspProThr 527
          |||:::|||   :::||||||
7173 CCTCGCCGCGTGCGATGTCGCCGACC 7198
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2000, 20:00:03 ; Search time 962.55 Seconds
(without alignments)
8457.048 Million cell updates/sec

Title: US-09-304-121-1
Perfect score: 2156
Sequence: 1 cgggcccgttgcaagatggc.....aaaaaaaaaaaaaaaaaaaa 2156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	697.8	32.4	763	49	AI628965	AI628965 ty79a02.x
2	697.4	32.3	730	64	AW054829	AW054829 ws60d05.x

	3	692.2	32.1	795	45	AI393937	AI393937	tg11a08.x
	4	686.2	31.8	751	49	AI651222	AI651222	wa98b09.x
	5	681.2	31.6	715	63	AW007349	AW007349	ws51d12.x
	6	664	30.8	701	61	AI809542	AI809542	wf30g01.x
	7	643.6	29.9	707	49	AI634255	AI634255	tr84d01.x
	8	627.6	29.1	927	44	AI324484	AI324484	mo99b07.y
	9	604.6	28.0	631	50	AI700961	AI700961	we09b01.x
c	10	585.2	27.1	607	61	AI810657	AI810657	tu19e02.x
c	11	583.6	27.1	961	44	AI325062	AI325062	mo99b07.x
c	12	578.6	26.8	610	63	AI937241	AI937241	wp74g09.x
c	13	555.4	25.8	776	42	AI093471	AI093471	qb15a03.x
c	14	544	25.2	544	41	AI041216	AI041216	ov77h08.x
	15	538.4	25.0	599	47	AI521804	AI521804	ti82f04.x
	16	521	24.2	524	73	AW169960	AW169960	xj35b05.x
c	17	520.6	24.1	670	62	AI884892	AI884892	wl87a10.x
	18	508.4	23.6	522	61	AI863994	AI863994	wj54e02.x
c	19	502.8	23.3	522	61	AI831101	AI831101	wj62h07.x
	20	501	23.2	547	50	AI692361	AI692361	wd63e02.x
c	21	486.2	22.6	544	41	AI026864	AI026864	ov88d12.x
	22	485	22.5	485	50	AI703424	AI703424	we24a12.x
c	23	484.6	22.5	554	36	AA614127	AA614127	no82e11.s
	24	481.4	22.3	497	63	AI934773	AI934773	wp89c05.x
	25	471.6	21.9	537	31	AA291542	AA291542	zt41e12.s
	26	470.4	21.8	651	74	AW177014	AW177014	CM3-CT010
	27	470	21.8	481	41	AI042312	AI042312	oy13h07.x
	28	468.6	21.7	641	27	AA044583	AA044583	zk73g12.r
	29	462	21.4	517	24	N34260	N34260	yx79f03.r1
c	30	461	21.4	461	63	AI984135	AI984135	wu21b08.x
c	31	460.4	21.4	475	64	AW044426	AW044426	wy72f08.x
c	32	460.2	21.3	504	42	AI150385	AI150385	qf40g02.x
c	33	460	21.3	466	46	AI440385	AI440385	tc83c09.x
	34	455	21.1	463	69	AW137844	AW137844	UI-H-BI1-
c	35	454.4	21.1	537	37	AA722900	AA722900	zg89h11.s
	36	453	21.0	580	36	AA619537	AA619537	vo84h01.r
	37	452	21.0	461	69	AW135472	AW135472	UI-H-BI1-
c	38	451.4	20.9	519	38	AA757068	AA757068	ah55c03.s
	39	450.8	20.9	546	37	AA721970	AA721970	zh17b06.s
	40	450.2	20.9	492	43	AI191359	AI191359	qe32d05.s
	41	441.2	20.5	486	24	N34263	N34263	yx79f06.r1
c	42	438.2	20.3	499	43	AI184729	AI184729	qd64c08.x
	43	437.2	20.3	450	38	AA808226	AA808226	oc40f09.s
c	44	432.8	20.1	500	35	AA563861	AA563861	nk19b04.s
c	45	427	19.8	466	36	AA641538	AA641538	nr79a10.s

ALIGNMENTS

```
RESULT   1
AI628965
LOCUS       AI628965       763 bp    mRNA            EST       23-APR-1999
DEFINITION  ty79a02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2285258 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI628965
VERSION     AI628965.1  GI:4665765
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 763)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 16, 1998 this sequence version replaced gi:2961738.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .763
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2285258"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      175 a    236 c    219 g    123 t     10 others
ORIGIN
```

```
  Query Match             32.4%;  Score 697.8;  DB 49;  Length 763;
  Best Local Similarity   95.0%;  Pred. No. 5.3e-120;
  Matches  725;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy    163 ggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctgaccaa 222
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||| .
Db      1 GGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCATCAACGTCCCGGCCTTCCTGACCAA 60

Qy    223 gctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccgagcgg 282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGG 120

Qy    283 gaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgctgcccaagtactt 342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAACAGCTTCCACGTGTTCGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTT 180

Qy    343 caagcacaacaacatggccagcttcgtgcggcagctcaacatgtatggcttccggaaagt 402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGCACAACAACATGGCCAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGT 240

Qy    403 ggtccacatcgagcagggcggcctggtcaagccagagagagacgacacggagttccagca 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGTCCACATCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCA 300

Qy    463 cccatgcttcctgcgtggccaggagcagctccttgagaacatcaagaggaaagtgaccag 522
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAG 360

Qy    523 tgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgtcaccaagctgct 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCT 420

Qy    583 gacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaagctcctggccat 642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GACGGACGTGCAGCTGATGAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCAT 480

Qy    643 gaagcatgagaatgaggctctgtggcgggaggtggccagccttcggcagaagcatgccca 702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAAGCATGAGAATGAGGCTCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCA 540

Qy    703 gcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggtgcagtcaaaccg 762
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCAACAGAAAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCG 600
```

```
Qy    763 gatcctgggggtgaagagaaagatccccctgatgctgaacgacagtggctcagcacattc 822
          ||  | |    |    ||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATNCCCTGNGNGTGAAGAGAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTC 660

Qy    823 catgcccaagtat-agccggcagttctccctggagcacgtccacggctcgggcccctact 881
          ||||||| ||||| || |      |||||||||   |||| ||| | ||||||||||||
Db    661 CATGCCCCAGTATAAGNCCCGCAGTCTCCCTGGNAGACGTNCACCGGTCGGGCCCCTACT 720

Qy    882 cggccccctccccagcctacagcagctccagcctctacgcccc 924
          ||| ||||  |||||||||||||||   |||||| || |||||
Db    721 CGGNCCCCCTCCCAGCCTACAGCAGNNTCAGCCTNTAAGCCCC 763


RESULT  2
AW054829
LOCUS       AW054829      730 bp    mRNA            EST       23-SEP-1999
DEFINITION  ws60d05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2501577 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AW054829
VERSION     AW054829.1  GI:5920532
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187940.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2501577"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      167 a    215 c    223 g    125 t
ORIGIN


  Query Match             32.3%;  Score 697.4;  DB 64;  Length 730;
  Best Local Similarity   98.3%;  Pred. No. 6.2e-120;
  Matches  715;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy    124 gcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccccgg 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGG 60

Qy    184 cgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgagcga 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGA 120

Qy    244 cccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgttcga 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGA 180

Qy    304 ccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggccag 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAG 240

Qy    364 cttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagggcgg 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGG 300

Qy    424 cctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtggcca 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCA 360

Qy    484 ggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtga 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGA 420

Qy    544 agacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaa 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAA 480

Qy    604 ggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggctct 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCT 540

Qy    664 gtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaa 723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAACAA 600

Qy    724 gctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagagaaa 783
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCTCATTCAGTTGCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGAGAAA 660

Qy    784 gatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccggca 843
          |||||  |||||||||  ||||||||||||||||||||| |||||  || ||||||  |||||
Db    661 GATCCTGCTGATGCTGTACGACAGTGGCTCAGCACATGCCATG-CCGAGTATATGCGGCA 719

Qy    844 gttctcc 850
          ||   ||
Db    720 GTCTGCC 726


RESULT  3
AI393937
LOCUS       AI393937      795 bp    mRNA            EST       30-MAR-1999
DEFINITION  tg11a08.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2108438 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI393937
VERSION     AI393937.1  GI:4223484
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 795)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 5, 1998 this sequence version replaced gi:2747316.
            Contact: Robert Strausberg, Ph.D.
```

```
           Tel: (301) 496-1550
           Email: Robert_Strausberg@nih.gov
           Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
           M.D., Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html

           Insert Length: 878   Std Error: 0.00
           Seq primer: -40UP from Gibco
           High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .795
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2108438"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      176 a    249 c    234 g    136 t
ORIGIN


  Query Match             32.1%;  Score 692.2;  DB 45;  Length 795;
  Best Local Similarity   95.7%;  Pred. No. 5.8e-119;
  Matches  765;  Conservative    0;  Mismatches   28;  Indels    6;  Gaps    5;

Qy    124 gcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccccgg 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGG 60

Qy    184 cgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgagcga 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGA 120

Qy    244 cccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgttcga 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGA 180

Qy    304 ccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggccag 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAG 240

Qy    364 cttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagggcgg 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGG 300

Qy    424 cctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtggcca 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCA 360

Qy    484 ggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtga 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGA 420

Qy    544 agacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaa 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAA 480

Qy    604 ggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggctct 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCT 540

Qy    664 gtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaa 723
          ||||||||||| | |||||||| |||| |||||||||||||||   ||| |||  ||||
Db    541 GTGGCGGGAGGGTGGCAGCCTTCCGCAGGAGCATGCCCAGCAACCAGAAGGCGTGCACAA 600

Qy    724 gctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagagaaa 783
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    601 GCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCCGATCCTGGGGGTGAAGAGAAA 660

Qy    784 gatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccggca 843
          ||| |||||||||||||||||||||||| |||||||||||||||| ||||||||||||
Db    661 GAT-CCCCTGATGCTGAACGACAGTGGCT-AGCACATTCCATGCCCATGTATAGCCGGCA 718

Qy    844 gttctccctggagcacgtccacggctcgggcccctactcggcccccctccccagcctacag 903
          | ||||| |||||||||||||||| || |||||||||||| || |  |||    |||  |
Db    719 G-TCTCCTTGGAGCACGTCCACGGATC-GGCCCCTACTCGTCCACTGCCC--AGCTAACG 774

Qy    904 cagctccagcctctacgcc 922
          |||||||||| |   | ||
Db    775 CAGCTCCAGCTTTAGCCCC 793


RESULT   4
AI651222
LOCUS       AI651222      751 bp    mRNA            EST       04-MAY-1999
DEFINITION  wa98b09.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2304185 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI651222
VERSION     AI651222.1  GI:4735201
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 20, 1998 this sequence version replaced gi:2979980.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 483.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2304185"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
```

Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 170 a 228 c 222 g 125 t 6 others
ORIGIN

Query Match 31.8%; Score 686.2; DB 49; Length 751;
Best Local Similarity 95.9%; Pred. No. 7.4e-118;
Matches 722; Conservative 0; Mismatches 29; Indels 2; Gaps 2;

```
Qy    124 gcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccccgg 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGG 60

Qy    184 cgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgagcga 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGA 120

Qy    244 cccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgttcga 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGA 180

Qy    304 ccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggccag 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAG 240

Qy    364 cttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagggcgg 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGG 300

Qy    424 cctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtggcca 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCA 360

Qy    484 ggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtga 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGA 420

Qy    544 agacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaa 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAA 480

Qy    604 ggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggctct 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCT 540

Qy    664 gtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaa 723
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    541 GTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGANAGTCGTCAACAA 600

Qy    724 gctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagagaaa 783
          |||||||||| |||||||||||||    ||||||   |  |||| |||||| ||
Db    601 GCTCATTCAGTNTCTGATCTCACTGGTGCGTCAAAACCGATNNCCTGGGGGGAAGAG-AA 659

Qy    784 gatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccggca 843
          |||||||||||||||||  |||||||||||||||||||||||| |||     || | |||
Db    660 GATCCCCCTGATGCTGAAACACAGTGGCTCAGCACATTCCATGGCCAGTATAAGGCCGCA 719

Qy    844 gttctccctggagcacgtccacggctcgggccc 876
           ||||||||||   |||| |||||||||| |||
Db    720 -TTCTCCCTGGNAGACGTNCACGGCTCGGCCCC 751
```

RESULT 5
AW007349
LOCUS AW007349 715 bp mRNA EST 10-SEP-1999
DEFINITION ws51d12.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2500727 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA sequence.
ACCESSION AW007349
VERSION AW007349.1 GI:5856127
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 715)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188853.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 489.
FEATURES Location/Qualifiers
source 1. .715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2500727"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 168 a 213 c 212 g 121 t 1 others
ORIGIN

Query Match 31.6%; Score 681.2; DB 63; Length 715;
Best Local Similarity 97.9%; Pred. No. 6.2e-117;
Matches 700; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```
Qy    124 gcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccccgg 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGG 60

Qy    184 cgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgagcga 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGA 120

Qy    244 cccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgttcga 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGA 180

Qy    304 ccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggccag 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAG 240

Qy    364 cttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagggcgg 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGG 300

Qy    424 cctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtggcca 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCA 360
```

```
Qy    484 ggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtga 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGA 420

Qy    544 agacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaa 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAA 480

Qy    604 ggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggctct 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCT 540

Qy    664 gtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaa 723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAACAA 600

Qy    724 gctcattcagttcctgatctcactggtgcagtc-aaaccggatcctgggggtgaagagaa 782
          ||||| |||| | |||||||||||||||||||| ||||||  |||||||||||||||  |
Db    601 GCTCAATCAGNTTCTGATCTCACTGGTGCAGTCAAAACCGATTCCTGGGGGTGAAGAAGA 660

Qy    783 agatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatag 837
          |||||||||||||||||  ||||||||||||||| ||||||   | |||||||
Db    661 AGATCCCCCTGATGCTGAAACACAGTGGCTCAGCACCTTCCATTGGCCAGTATAG 715
```

RESULT 6
AI809542
LOCUS AI809542 701 bp mRNA EST 07-JUL-1999
DEFINITION wf30g01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2357136 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA sequence.
ACCESSION AI809542
VERSION AI809542.1 GI:5396108
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 701)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247218.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 480.
FEATURES Location/Qualifiers
source 1. .701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2357136"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 162 a 209 c 211 g 118 t 1 others
ORIGIN

```
Query Match            30.8%;  Score 664;  DB 61;  Length 701;
Best Local Similarity  97.6%;  Pred. No. 9.6e-114;
Matches  684;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy    124 gcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccccgg 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGG 60

Qy    184 cgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgagcga 243
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGCGGGGGCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGA 120

Qy    244 cccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgttcga 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGA 180

Qy    304 ccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggccag 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAG 240

Qy    364 cttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagggcgg 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGG 300

Qy    424 cctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtggcca 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCA 360

Qy    484 ggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtga 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGA 420

Qy    544 agacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaa 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAA 480

Qy    604 ggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggctct 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCT 540

Qy    664 gtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaa 723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAACAA 600

Qy    724 gctcattcagttcctgatctcactggtgcagtc-aaaccggatcctgggggtgaagagaa 782
          ||||||| | |||||||||||||||||||| |||||||||  |||| | |
Db    601 GCTCATTTCAGTTCTGATCTCACTGGTGCAGTCAAAACCGGATCNCTGGGGGTGAAGAGA 660

Qy    783 agatccccctgatgctgaacgacagtggctcagcacattcc 823
          |||||||||||||||||||||||||||||||||||||||||
Db    661 AGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCC 701
```

RESULT 7
AI634255
LOCUS AI634255 707 bp mRNA EST 26-APR-1999
DEFINITION tr84d01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224993 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA sequence.
ACCESSION AI634255
VERSION AI634255.1 GI:4685585
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 707)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3138329.

```
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2224993"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT      169 a    207 c    211 g    116 t      4 others
ORIGIN


  Query Match             29.9%;  Score 643.6;  DB 49;  Length 707;
  Best Local Similarity   95.2%;  Pred. No. 5.8e-110;
  Matches  673;  Conservative    0;  Mismatches   32;  Indels    2;  Gaps    1;

Qy    160 gatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctgac 219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGAC 60

Qy    220 caagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccgag 279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAG 120

Qy    280 cgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgctgcccaagta 339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGGAACAGCTTCCACGTGTTCGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTA 180

Qy    340 cttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtatggcttccggaa 399
          |||||||||||||||||||||||||||||||||||||||||||||||||| | |||||||
Db    181 CTTCAAGCACAACAACATGGCCAGCTTCGTGCGGCAGCTCAACATGTATGGCTTTCCGAA 240

Qy    400 agtggtccacatcgagcagggcggcctggtcaagccagagagagacgacacggagttcca 459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTGGTCCACATCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCA 300

Qy    460 gcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaagaggaaagtgac 519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGAC 360

Qy    520 cagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagcgtcaccaagct 579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCT 420

Qy    580 gctgacggacgtgcagctgatgaaggggaagcaggagtgcatggactccaagctcctggc 639
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    421 GCTGACGGACGTGCAGCTGATGAAGGGGAAGCAGGAGTGCATGGACTNCAAGCTCCTGGC 480

Qy    640 catgaagcatgagaatgaggctctgtggcgggaggtggccagccttcggcagaagcatgc 699
          |||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||||
Db    481 CATGAAGCATGAGAATGAGGCTCTGTGGCGGGAGGGTGGCAGCCTTCGGCAGAAGCATGC 540

Qy    700 ccagcaacagaaagtcgtcaacaagctcattcagttcctgatctcactggtgcagtcaaa 759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCAGCACCAGAAAGTCGTCAACAAGCTCATTCAGGTCCTGATCTCACTGGTGCAGTCAAA 600

Qy    760 ccggatcctgggggtgaagagaaagatccccctgatgctgaacgacagtggctcagcaca 819
          | |   |   |||   |    ||||||||||||||||||||||||||||||||||||||
Db    601 ACCGGATCCCTGGGCTGAAGAGAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACA 660

Qy    820 ttccatgcccaagtat--agccggcagtttctccctggagcacgtcca 864
          |||||||||| |||||  || ||    |  |||||||||||||| ||
Db    661 TTCCATGCCCCAGTATANAGACGCAGNTCTTCCCTGGAGCACGTNCA 707


RESULT   8
AI324484
LOCUS       AI324484      927 bp    mRNA            EST       23-DEC-1998
DEFINITION  mo99b07.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:567829 5' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1
            (HUMAN); gb:X61753 M.musculus mRNA for heat shock transcription
            factor 1 (MOUSE);, mRNA sequence.
ACCESSION   AI324484
VERSION     AI324484.1  GI:4058913
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152840.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:342477
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:567829"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      225 a    262 c    241 g    191 t      8 others
ORIGIN


  Query Match             29.1%;  Score 627.6;  DB 44;  Length 927;
  Best Local Similarity   85.1%;  Pred. No. 5.7e-107;
  Matches  766;  Conservative    0;  Mismatches  127;  Indels    7;  Gaps    6;

Qy    240 gcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgt 299
          ||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db     10 GCGACCCGGACACAGACGCGCTCATCTGCTGGAGCCCGAGTGGGAACAGCTTCCACGTGT 69
```

```
Qy    300 tcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatgg 359
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 TTGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGG 129

Qy    360 ccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagg 419
          | |||||||||||||||||||||||||||||||| ||||| |||||||| |||||||
Db    130 CTAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGAAAAGTAGTCCACATTGAGCAGG 189

Qy    420 gcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtg 479
          | ||||||||||||||| |||||||| ||||| |||||||||||| || || ||| ||||||
Db    190 GTGGCCTGGTCAAGCCTGAGAGAGATGACACCGAGTTCCAGCATCCTTGTTTCTTGCGTG 249

Qy    480 gccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaaga 539
          | ||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    250 GACAGGAACAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGCGTGTCCACCCTGAAGA 309

Qy    540 gtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctga 599
          |||| |||||||| || |||||||||| ||||||  |||| |||| || ||||||||||
Db    310 GTGAGGACATAAAAATACGCCAGGACAGTGTCACCCGGCTGTTGACAGATGTGCAGCTGA 369

Qy    600 tgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgagg 659
          |||||||||| |||||||| |||||||||||||| |||||||||||||| ||||| ||||
Db    370 TGAAGGGGAAACAGGAGTGTATGGACTCCAAGCTNCTGGCCATGAAGCACGAGAACGAGG 429

Qy    660 ctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtca 719
          | ||||||| |||||  ||||||| |||||||||||||||||||| || ||||| ||||
Db    430 CCCTGTGGC-GGAGGGTGCCAGCCCTCGGCAGAAGCATGCCCAGCAGCAAAAAGTTGTCA 488

Qy    720 acaagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaaga 779
          |||||||||||||||||||||||||||||||||||| ||||||||||| |||||||||
Db    489 ACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCGAACCGGATCCTGNGGGTGAAGA 548

Qy    780 gaaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagcc 839
          |||||||||| |||||| || ||||| |||||||||| || ||||||||||| | |
Db    549 GAAAGATCCCTCTGATGTNGAGTGACAGCAACTCAGCACACTCTGTGCCCAAGTATGGTC 608

Qy    840 ggcagttctccctggagcacgtccacggct-cgggcccctactcggcccctccccagcc 898
          | |||| || |||||||||| || || || | | ||||| ||||| || || ||  |||||
Db    609 GACAGTACTTCCTGGAGCATGTNCATGGGTCCTGGCCCATACTCAGCTCCATCTTCAGCC 668

Qy    899 tacagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctcc 958
          |||||||||||| ||||| ||| | ||||||||||  |||||||| ||||||| |||| |
Db    669 TACAGCAGCTCTAGCCTTTACTTCTCTGATGCTGTCACCAGCTCTTGACCCATAATCTNC 728

Qy    959 gacatcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacgag 1018
          || ||||| |||||||||||  |    ||| | | |  ||| || |||||||||| |||
Db    729 GATATCACTGAGCTGGCTCCAAC--AGCCCTTTGGCCTCCCTAGCAGGAGCATAGATGAG 786

Qy   1019 aggcccctatccagcagcccccctggtgcgtgtcaaggaggagccccccagcccgcctcag 1078
          | ||| || | ||||||| | |||   |||||| || | |||||||||| |  ||| |
Db    787 AAGCCTCTGTNCAGCAGCACTCTGTCCCGTGTC-AGCAAGAGCCCCCCAGNCAACCT-AC 844

Qy   1079 agcccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtcc 1138
          || ||   ||||  |||||| || ||| | || ||||| | | ||   ||||| | | ||
Db    845 AGGCCTTCGGTACTGGAGGCAAG-CCCTGCCGGCCATCCTNCATGATAACCCTTTGGCCC 903


RESULT   9
AI700961
LOCUS       AI700961      631 bp    mRNA            EST       03-JUN-1999
DEFINITION  we09b01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2340553 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI700961
VERSION     AI700961.1  GI:4988861
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137377.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2340553"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      147 a    184 c    191 g    108 t      1 others
ORIGIN


  Query Match             28.0%;  Score 604.6;  DB 50;  Length 631;
  Best Local Similarity   98.3%;  Pred. No. 9.6e-103;
  Matches  621;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy    142 tattccctccttgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaa 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAA 60

Qy    202 cgtcccggccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgct 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGCT 120

Qy    262 catctgctggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaa 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGCCAGTTTGCCAA 180

Qy    322 ggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaa 381
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    181 GGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAGCTTCGTGCGGCAGCTCAA 240

Qy    382 catgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagag 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGGCCTGGTCAAGCCAGAGAG 300

Qy    442 agacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaa 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAA 360

Qy    502 catcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgcca 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCCA 420

Qy    562 ggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcat 621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    421 GGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAGCAGGAGTGCAT 480

Qy    622 ggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccag 681
          ||||||||||||||||||||||||||||||||||||||||   |  ||||||||||||
Db    481 GGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCTNGTGGCGGAGGTGGCCAG 540

Qy    682 ccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgat 741
          |||||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db    541 CCTTCGGCAGAAGCATGCTCAGCAACAGAAAGTCGTCAACAAGCTCATTCAGTTTCTGAT 600

Qy    742 ctcactggtgcagtcaaaccggatcctggggg 773
          |||||||||| ||||||||||   ||||||||
Db    601 CTCACTGGTG-AGTCAAACCGATCCCTGGGGG 631


RESULT 10
AI810657/c
LOCUS       AI810657      607 bp    mRNA            EST       07-JUL-1999
DEFINITION  tu19e02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251514 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI810657
VERSION     AI810657.1  GI:5397223
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948429.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251514"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      113 a    178 c    191 g    124 t      1 others
ORIGIN


  Query Match           27.1%;  Score 585.2;  DB 61;  Length 607;
  Best Local Similarity  99.3%;  Pred. No. 3.7e-99;
  Matches  587;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   1556 gtgcactacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagc 1615
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    591 GTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCGGGCTCCGTGGACACCGGGAGC 532

Qy   1616 aacgacctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggc 1675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 AACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGC 472

Qy   1676 ttcgccgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggac 1735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 TTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGAC 412

Qy   1736 cccactgtctcctagaggccccggaggagctgggccagccgcccacccccacccccagtg 1795
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    411 CCCACTGTCTCCTAGAGGCCCCGGAGGAGNTGGGCCAGCCGCCCACCCCCACCCCCAGTG 352

Qy   1796 cagggctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccg 1855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 CAGGGCTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCG 292

Qy   1856 ccatagccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtc 1915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 CCATAGCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTC 232

Qy   1916 accctggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttc 1975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 ACCCTGGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTC 172

Qy   1976 acagccacacctggactgaccctgcaggttgttcatagtcagaattgtattttggatttt 2035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 ACAGCCACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTT 112

Qy   2036 tacacaactgtcccgttccccgctccacagagatacacagatatatacacacagtggatg 2095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 TACACAACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATG 52

Qy   2096 gacggacaagacaggcagagatctataaacagacaggctctaaaaaaaaaa 2146
          |||||||||||||||||||||||||||||||||||||||||  |||||||
Db     51 GACGGACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTATGAAAAAAA 1


RESULT 11
AI325062/c
LOCUS       AI325062      961 bp    mRNA            EST       23-DEC-1998
DEFINITION  mo99b07.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:567829 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1
            (HUMAN); gb:X61753 M.musculus mRNA for heat shock transcription
            factor 1 (MOUSE);, mRNA sequence.
ACCESSION   AI325062
VERSION     AI325062.1  GI:4059491
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 961)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151652.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
```

```
          IMAGE Consortium (info@image.llnl.gov) for further information.
          MGI:342477
          This clone was previously sequenced on the 5' end only, this new
          data is from the 3' end
          High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .961
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:567829"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      190 a    227 c    314 g    221 t      9 others
ORIGIN


  Query Match             27.1%;  Score 583.6;  DB 44;  Length 961;
  Best Local Similarity   80.8%;  Pred. No. 8.2e-99;
  Matches  728;  Conservative    0;  Mismatches  158;  Indels   15;  Gaps    4;

Qy     551 aagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaag 610
           || || ||||||||||||  |||||  |||| ||   || |||||| | |||||||| ||
Db     958 AAAATACGCCAGGACAGTTTCACCGGGCTGTTGCAAGATGTGCAG-TTATGAAGGGAAAC 900

Qy     611 caggagtgcatggactccaagc-tcctggccatgaagcatgagaatgaggctctgtggcg 669
           ||| |||| ||||||| ||||| || ||| |||||||| |||||  |||  || ||||||
Db     899 CAGAAGTGTATGGACTTCAAGCTTCNTGGACATGAAGCACGAGAACAAGGGCCTTTGGCG 840

Qy     670 ggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcat 729
           | ||||||| ||||| ||||||||||||||    || || ||||| ||||||||||||||
Db     839 GAAGGTGGCAAGCCTCCGGCAGAAGCATGCCAGCCAGCAAAAAGTTGTCAACAAGCTCAT 780

Qy     730 tcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagagaaagatccc 789
           |||||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db     779 TCAGTTCCTGATCTCACTGGTGCAGTCGAACCGGATTCTGGGGGTGAAGAGAAAGATCCT 720

Qy     790 cctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccggcagttctc 849
           |||||| |||  |||||   |||||| || ||  |||||||||| | || |||| |||
Db     719 TCTGATGTTGAGTGACAGCAACTCAGCNCACTCTGTGCCCAAGTATGGTCGACAGTACTC 660

Qy     850 cctggagcacgtccacggctcgggcccctactcggccccctccccagcctacagcagctc 909
           |    |||| ||||| ||    ||||| ||||| || || || ||||| ||||||||||
Db     659 CTGNNAGCATGTCCATGGTCNTGGCCCATACTCAGCTCCATCTCCAGCNTACAGCAGCTC 600

Qy     910 cagcctctacgcccctgatgctgtggccagctctggacccatcatctccgacatcaccga 969
           |||  | ||| || ||||||||||  |||||||||||||||| |||||||| ||||| ||
Db     599 TAGCATTTACTCCTCTGATGCTGTCACCAGCTCTGGACCCATAATCTCCGATATCACTGA 540

Qy     970 gctggctcctgccagccccatggcctcccccggcgggagcatagacgagaggcccctatc 1029
           ||||||||   |||||||  |||||||| || ||| |||||||||| ||||||| || ||
Db     539 GCTGGCTCNCACCAGCCCTNTGGCCTCNCCAGGCAGGAGCATAGATGAGAGGCCTCTGTC 480

Qy    1030 cagcagcccccctggtgcgtgtcaaggaggagccccccagcccgcctcagagcccccgggt 1089
           ||||||| | ||||| ||||||||| | |||||| ||||||| ||||| ||||| |||||
Db     479 CAGCAGCACTCTGGT-CGTGTCAAGCAAGAGCCCACCAGCCCACCTCACAGCCCTCGGGT 421

Qy    1090 agaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtccccgaccgccct 1149
           |  ||||||||| || |||||||||| ||| |||| ||||  |||||||| || ||| |
Db     420 ACTGGAGGCGAGCCCTGGGCGCCCATCCTCCATGGATACCCCTTTGTCCCCAACTGCCTT 361

Qy    1150 cattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagccctcacgga 1209
           |||||||||||||| || ||||| || ||| |||| ||  ||   |   |   | |
Db     360 CATTGACTCCATCCTTCGAGAGAGCGAGCCTACCCCTGCTGCCTCAAACACAGCCCCTAT 301

Qy    1210 cgccaggggccacacggacaccgagggccggcctccctccccccgcccacctccacccc 1269
           | ||              ||||  | ||   | ||| | | |||| || ||||||||||
Db     300 GGACA-----------CAACCGGAGCCCAAGCCCCCGCACTCCCGACCCCCTCCACCCC 253

Qy    1270 tgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgaccacttggatgc 1329
           ||| |||||||||||||||||||||| |||||||| ||||| ||||| ||| |||||||
Db     252 TGAGAAGTGCCTCAGCGTAGCCTGCCTAGACAAGAACGAGCTAAGTGATCACCTGGATGC 193

Qy    1330 tatggactccaacctggataacctgcagaccatgctgagcagccacggcttcagcgtgga 1389
           |||||||||||||||| ||||||||||||||||||  |||||||||||||| |||||
Db     192 CATGGACTCCAACCTGGACAACCTGCAGACCATGCTGACAAGCCACGGCTTCAGTGTGGA 133

Qy    1390 caccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacatgagcctgcc 1449
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     132 CACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCATGCCCGACATGAGCCTGCC 73

Qy    1450 t 1450
           |
Db      72 T 72


RESULT  12
AI937241/c
LOCUS       AI937241      610 bp    mRNA            EST       06-SEP-1999
DEFINITION  wp74g09.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2467552 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI937241
VERSION     AI937241.1  GI:5676111
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187628.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2467552"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
```

Soares and M.Fatima Bonaldo."
BASE COUNT 112 a 179 c 194 g 123 t 2 others
ORIGIN

Query Match 26.8%; Score 578.6; DB 63; Length 610;
Best Local Similarity 98.9%; Pred. No. 6.3e-98;
Matches 603; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```
Qy   1539 attcagggaagcagctggtgcactacacagcgcagccg-ctgttcctgctggaccccggc 1597
          |||||||||||||||||||||||||||||||||||| |||||||||||||||| |||
Db    610 ATTCAGGGAAGCAGCTGGTGCACTACACAGCGCAGCCGNCTGTTCCTGCTGGACCCGGGC 551

Qy   1598 tccgtggaca-ccgggagcaacgacctgccggtgctgtttgagctgggagagggctccta 1656
          |||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||
Db    550 TCCGTGGACACCCGGGAGCAACGACCTGCCGGTGCTGTNTGAGCTGGGAGAGGGCTCCTA 491

Qy   1657 cttctccgaaggggacggcttcgccgaggaccccaccatctccctgctgacaggctcgga 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 CTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGA 431

Qy   1717 gcctcccaaagccaaggaccccactgtctcctagaggccccggaggagctgggccagccg 1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 GCCTCCCAAAGCCAAGGACCCCACTGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCG 371

Qy   1777 cccacccccaccccccagtgcagggctggtcttggggaggcagggcagcctcgcggtcttg 1836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 CCCACCCCCACCCCCAGTGCAGGGCTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTG 311

Qy   1837 ggcactggtgggtcggccgccatagccccagtaggacaaacgggctcgggtctgggcagc 1896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 GGCACTGGTGGGTCGGCCGCCATAGCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGC 251

Qy   1897 acctctggtcaggagggtcaccctggcctgccagtctgccttccccccaaccccgtgtcct 1956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 ACCTCTGGTCAGGAGGGTCACCCTGGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCT 191

Qy   1957 gtggtttggttggggcttcacagccacacctggactgaccctgcaggttgttcatagtca 2016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 GTGGTTTGGTTGGGGCTTCACAGCCACACCTGGACTGACCCTGCAGGTTGTTCATAGTCA 131

Qy   2017 gaattgtattttggatttttacacaactgtcccgttccccgctccacagagatacacaga 2076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 GAATTGTATTTTGGATTTTTACACAACTGTCCCGTTCCCCGCTCCACAGAGATACACAGA 71

Qy   2077 tatatacacacagtggatggacggacaagacaggcagagatctataaacagacaggctct 2136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 TATATACACACAGTGGATGGACGGACAAGACAGGCAGAGATCTATAAACAGACAGGCTCT 11

Qy   2137 aaaaaaaaaa 2146
          |   ||||||
Db     10 ATGCAAAAAA 1
```

RESULT 13
AI093471/c
LOCUS AI093471 776 bp mRNA EST 18-AUG-1998
DEFINITION qb15a03.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:1696300 3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);contains TAR1.t1 PTR5 repetitive element ;, mRNA sequence.
ACCESSION AI093471
VERSION AI093471.1 GI:3432447
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 776)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151925.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 458.
FEATURES Location/Qualifiers
source 1. .776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1696300"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
BASE COUNT 133 a 222 c 257 g 162 t 2 others
ORIGIN

Query Match 25.8%; Score 555.4; DB 42; Length 776;
Best Local Similarity 99.8%; Pred. No. 1.3e-93;
Matches 556; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   1580 ttcctgctggaccccggctccgtggacaccgggagcaacgacctgccggtgctgtttgag 1639
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 TCCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGACCTGCCGGTGCTGTTTGAG 498

Qy   1640 ctgggagagggctcctacttctccgaaggggacggcttcgccgaggaccccaccatctcc 1699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 CTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCC 438

Qy   1700 ctgctgacaggctcggagcctcccaaagccaaggaccccactgtctcctagaggccccgg 1759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 CTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCTAGAGGCCCCGG 378

Qy   1760 aggagctgggccagccgcccaccccccaccccccagtgcagggctggtcttggggaggcagg 1819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 AGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGGCTGGTCTTGGGGAGGCAGG 318

Qy   1820 gcagcctcgcggtcttgggcactggtgggtcggccgccatagccccagtaggacaaacgg 1879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 GCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATAGCCCCAGTAGGACAAACGG 258

Qy   1880 gctcgggtctgggcagcacctctggtcaggagggtcaccctggcctgccagtctgccttc 1939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 GCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCCTGGCCTGCCAGTCTGCCTTC 198

Qy   1940 ccccaaccccgtgtcctgtggtttggttggggcttcacagccacacctggactgaccctg 1999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 CCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAGCCACACCTGGACTGACCCTG 138

Qy   2000 caggttgttcatagtcagaattgtattttggatttttacacaactgtcccgttccccgct 2059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 CAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACACAACTGTCCCGTTCCCCGCT 78

Qy   2060 ccacagagatacacagatatatacacacagtggatggacggacaagacaggcagagatct 2119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 CCACAGAGATACACAGATATATACACACAGTGGATGGACGGACAAGACAGGCAGAGATCT 18

Qy   2120 ataaacagacaggctct 2136
          |||||||||||||||||
Db     17 ATAAACAGACAGGCTCT 1
```

```
RESULT 14
AI041216/c
LOCUS       AI041216      544 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov77h08.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1643391
            3' similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1
            (HUMAN);contains MER22.t1 MER22 MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AI041216
VERSION     AI041216.1  GI:3280410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2045449.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 784   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1643391"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      102 a    160 c    174 g    108 t
ORIGIN


  Query Match           25.2%;  Score 544;  DB 41;  Length 544;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-91;
  Matches  544;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1596 gctccgtggacaccgggagcaacgacctgccggtgctgtttgagctgggagagggctcct 1655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 GCTCCGTGGACACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCT 485

Qy   1656 acttctccgaaggggacggcttcgccgaggaccccaccatctccctgctgacaggctcgg 1715
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 ACTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGG 425

Qy   1716 agcctcccaaagccaaggaccccactgtctcctagaggccccggaggagctgggccagcc 1775
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 AGCCTCCCAAAGCCAAGGACCCCACTGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCC 365

Qy   1776 gcccacccccacccccagtgcagggctggtcttggggaggcagggcagcctcgcggtctt 1835
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 GCCCACCCCCACCCCCAGTGCAGGGCTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTT 305

Qy   1836 gggcactggtgggtcggccgccatagccccagtaggacaaacgggctcgggtctgggcag 1895
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GGGCACTGGTGGGTCGGCCGCCATAGCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAG 245

Qy   1896 cacctctggtcaggagggtcaccctggcctgccagtctgccttccccccaaccccgtgtcc 1955
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 CACCTCTGGTCAGGAGGGTCACCCTGGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCC 185

Qy   1956 tgtggtttggttggggcttcacagccacacctggactgaccctgcaggttgttcatagtc 2015
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 TGTGGTTTGGTTGGGGCTTCACAGCCACACCTGGACTGACCCTGCAGGTTGTTCATAGTC 125

Qy   2016 agaattgtattttggatttttacacaactgtcccgttccccgctccacagagatacacag 2075
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 AGAATTGTATTTTGGATTTTTACACAACTGTCCCGTTCCCCGCTCCACAGAGATACACAG 65

Qy   2076 atatatacacacagtggatggacggacaagacaggcagagatctataaacagacaggctc 2135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 ATATATACACACAGTGGATGGACGGACAAGACAGGCAGAGATCTATAAACAGACAGGCTC 5

Qy   2136 taaa 2139
           ||||
Db      4 TAAA 1


RESULT 15
AI521804
LOCUS       AI521804      599 bp    mRNA            EST       13-APR-1999
DEFINITION  ti82f04.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2138527 3'
            similar to gb:M64673 HEAT SHOCK FACTOR PROTEIN 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AI521804
VERSION     AI521804.1  GI:4435939
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948550.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 893   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2138527"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
```

Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 138 a 178 c 181 g 101 t 1 others
ORIGIN

Query Match 25.0%; Score 538.4; DB 47; Length 599;
Best Local Similarity 99.5%; Pred. No. 1.8e-90;
Matches 561; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
Qy    142 tattccctccttgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaa 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAA 60

Qy    202 cgtcccggccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgct 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGCT 120

Qy    262 catctgctggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaa 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGCCAGTTTGCCAA 180

Qy    322 ggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaa 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGCCAGCTTCGTGCGGCAGCTCAA 240

Qy    382 catgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagag 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGGCCTGGTCAAGCCAGAGAG 300

Qy    442 agacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaa 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAA 360

Qy    502 catcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgcca 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCCA 420

Qy    562 ggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcat 621
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    421 GGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAGCATGAGTGCAT 480

Qy    622 ggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccag 681
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    481 GGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGCTCTGTGGCGGGA-GTGGCCAG 539

Qy    682 ccttcggcagaagcatgcccagca 705
          |||||||||| ||||||||||||
Db    540 CCTTCGGCAG-AGCATGCCCAGCA 562
```

Search completed: March 6, 2000, 20:18:14
Job time: 1091 sec

THIS PAGE BLANK (USPTO)

OM of: US-09-304-121-2 to: GenEmbl:* out_format : pfs

Date: Mar 7, 2000 2:15 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09304121/runat_06032000_151317_15712/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09304121
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-304-121-2
Query length: 529
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 844.650000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat:AR009723	+	2729.00	2470.40	1.4e-129	2156	! AR009723 Sequence 31 from pate
gb_pri:HUMHSF1	+	2729.00	2470.40	1.4e-129	2156	! M64673 Human heat shock factor
gb_ro:MMHSF1	+	2294.00	2077.72	1.0e-107	1947	! X61753 M.musculus mRNA for hea
gb_ro:RRHSF1	+	1999.00	1812.24	6.4e-93	1647	! X83094 R.rattus mRNA for heat
gb_ov:CHKHSF3A	+	1982.50	1796.69	4.7e-92	1774	! L06098 Chicken mRNA sequence.
gb_htg6:AF205589	-	1538.50	1359.02	1.1e-67	131973	! AF205589 Homo sapiens chromo
gb_ov:XELHSF	+	1406.50	1276.17	4.6e-63	1657	! L36924 Xenopus laevis heat sho
gb_ro:AF059275	+	1393.00	1247.85	1.8e-61	11395	! AF059275 Mus musculus heat sh
gb_ov:CHKHSF3B	+	814.00	737.18	4.9e-33	2366	! L06125 Chicken mRNA sequence.
gb_pri:HUMHSF2	+	791.50	716.67	6.8e-32	2411	! M65217 Human heat shock factor
gb_ro:AF172640	+	787.00	715.08	8.3e-32	1792	! AF172640 Rattus norvegicus hea
gb_ro:MMHSF2	+	774.00	702.51	4.2e-31	1972	! X61754 M.musculus mRNA for hea
gb_pr2:AB029348	+	763.00	694.43	1.2e-30	1577	! AB029348 Homo sapiens mRNA for
gb_ro:AB029349	+	762.00	692.97	1.4e-30	1686	! AB029349 Mus musculus mRNA for
gb_ov:CHKHSF3C	+	740.00	669.21	3.0e-29	2675	! L06126 Chicken mRNA sequence.
gb_pri:D87673	+	717.00	652.93	2.4e-28	1555	! D87673 Homo sapiens mRNA for H
gb_ro:AB029350	+	717.00	652.74	2.5e-28	1591	! AB029350 Mus musculus mRNA for
gb_in1:DROHSPHEX	+	705.50	637.74	1.7e-27	2760	! M60070 D.melanogaster heat sho
gb_pat:AR009733	+	705.50	637.67	1.7e-27	2781	! AR009733 Sequence 42 from pate
gb_in2:AC004336	+	502.00	427.74	8.4e-16	61267	! AC004336 Drosophila melanogas
gb_in2:AF043418	+	413.50	374.85	7.4e-13	2369	! AF043418 Schistosoma mansoni h
gb_in2:AF043416	+	413.00	377.89	5.0e-13	1560	! AF043416 Schistosoma mansoni H
gb_pl1:YSPHSF	+	399.50	362.61	3.6e-12	2251	! M94683 Schizosaccharomyces pom
gb_in2:AF043417	+	396.50	359.86	5.1e-12	2261	! AF043417 Schistosoma mansoni h
gb_pr2:AB029347	+	393.00	349.70	1.9e-11	5221	! AB029347 Homo sapiens gene for
gb_pl1:KLHSFG	+	389.00	349.81	1.8e-11	3340	! X55149 K. lactis HSF gene for
gb_pl1:YSCHSF	+	362.00	324.50	4.7e-10	3713	! J03139 S.cerevisiae heat shock
gb_pl1:SCYGL074C	+	362.00	324.04	5.0e-10	3923	! Z72596 S.cerevisiae chromosome
gb_pl1:YSCHSF1A	+	360.00	322.23	6.3e-10	3926	! M22040 S.cerevisiae heat shock
gb_pl1:LPHSF8	+	346.00	316.12	1.4e-09	1791	! X67600 L.peruvianum Lp-hsf8 mR
gb_in2:SMHSF1	+	339.50	308.73	3.6e-09	2146	! AF043419 Schistosoma mansoni H
gb_pl2:SPAC2E12	+	328.50	286.91	5.9e-08	8884	! Z69726 S.pombe chromosome I co
gb_pl1:ATHSF3	+	327.00	298.93	1.3e-08	1791	! Y14068 Arabidopsis thaliana mR
gb_pl1:ATHSF1	+	323.50	290.61	3.6e-08	3319	! X76167 A.thaliana hsf1 gene. 3
gb_pl2:ATFCA9	-	321.50	262.03	1.4e-06	81835	! Z97344 Arabidopsis thaliana D
gb_pl2:ATU68561	+	299.00	275.62	2.5e-07	1407	! U68561 Arabidopsis thaliana he
gb_pl1:AB014484	+	295.50	270.92	4.6e-07	1690	! AB014484 Nicotiana tabacum NtH
gb_pl1:LPHSF30	+	294.00	271.66	4.1e-07	1315	! X67601 L.peruvianum Lp-hsf30 m
gb_pl1:GMHSF34	+	289.00	266.55	8.0e-07	1410	! Z46953 G.max mRNA for heat sho
gb_pl1:AB014483	+	274.50	252.25	5.0e-06	1625	! AB014483 Nicotiana tabacum NtH
gb_pl1:LPHSF24	+	267.00	245.97	1.1e-05	1530	! X55347 Tomato mRNA for heat st
gb_pl1:GMHSF21	+	265.50	251.50	5.5e-06	671	! Z46952 G.max mRNA for heat shod
gb_pl2:ATF13C5	-	265.00	207.78	0.0015	119111	! AL021711 Arabidopsis thal
gb_pl1:ATHSF4	+	259.50	241.61	2.0e-05	1144	! Y14069 Arabidopsis thaliana
gb_pl1:AB020746	+	258.50	205.25	0.0021	79706	! AB020746 Arabidopsis thali

seq_name: gb_pat:AR009723

seq_documentation_block:
LOCUS AR009723 2156 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 31 from patent US 5756343.
ACCESSION AR009723
VERSION AR009723.1 GI:3968528
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2156)
AUTHORS Wu,C., Clos,J., Westwood,J.Timothy and Rabindran,S.
TITLE Cell stress transcriptional factors
JOURNAL Patent: US 5756343-A 31 26-MAY-1998;
FEATURES Location/Qualifiers
source 1. .2156
/organism="unknown"
BASE COUNT 435 a 739 c 628 g 354 t
ORIGIN

alignment_scores:
Quality: 2729.00 Length: 529
Ratio: 5.159 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-304-121-2 x AR009723 ..

Align seg 1/1 to: AR009723 from: 1 to: 2156

```
      1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    161 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 210

     17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    211 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 260

     34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    261 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 310

     51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    311 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

     67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 410

     84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    411 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 460

    101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    461 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 510

    117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    511 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 560

    134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    561 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 610

    151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||
611 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 710

184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
711 AAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAAC 760

201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
761 CGGATCCTGGGGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTGG 810

217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
811 CTCAGCACATTCCATGCCCAAGTATAGCCGGCAGTTCTCCCTGGAGCACG 860

234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
861 TCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTCC 910

251 SerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAs 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
911 AGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

267 pIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerI 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCA 1010

284 leAspGluArgProLeuSerSerSerProLeuValArgValLysGluGlu 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1011 TAGACGAGAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAG 1060

301 ProProSerProProGlnSerProArgValGluGluAlaSerProGlyAr 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1061 CCCCCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGCG 1110

317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1111 CCCATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCCA 1160

334 leLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAsp 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1161 TCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGAC 1210

351 AlaArgGlyHisThrAspThrGluGlyArgProProSerProProProTh 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1211 GCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

367 rSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGluL 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGC 1310

384 euSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThr 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1311 TCAGTGACCACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACC 1360

401 MetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLe 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1361 ATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCT 1410

417 uPheSerProSerValThrValProAspMetSerLeuProAspLeuAspS 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1411 GTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACA 1460

434 erSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProArg 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1461 GCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGG 1510

451 ProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValHi 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
1511 CCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

 467 sTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThrG 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCG 1610

 484 lySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPhe 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1611 GGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTC 1660

 501 SerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1661 TCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGG 1710

 517 ySerGluProProLysAlaLysAspProThrValSer 529
     |||||||||||||||||||||||||||||||||||||
1711 CTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 1747
```

```
seq_name: gb_prl:HUMHSF1

seq_documentation_block:
LOCUS       HUMHSF1      2156 bp    mRNA            PRI       08-NOV-1994
DEFINITION  Human heat shock factor 1 (TCF5) mRNA, complete cds.
ACCESSION   M64673
VERSION     M64673.1  GI:184402
KEYWORDS    DNA-binding transcription factor; heat shock factor 1.
SOURCE      Homo sapiens lymphoid cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2156)
  AUTHORS   Rabindran,S.K., Giorgi,G., Clos,J. and Wu,C.
  TITLE     Molecular cloning and expression of a human heat shock factor, HSF1
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (16), 6906-6910 (1991)
  MEDLINE   91334376
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="B cell"
                     /tissue_type="lymphoid"
                     /map="20"
     gene            161. .1750
                     /gene="TCF5"
     CDS             161. .1750
                     /gene="TCF5"
                     /codon_start=1
                     /db_xref="GDB:G00-126-373"
                     /product="heat shock factor 1"
                     /protein_id="AAA52695.1"
                     /db_xref="GI:184403"
                     /translation="MDLPVGPGAAGPSNVPAFLTKLWTLVSDPDTDALICWSPSGNSF
                     HVFDQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVHIEQGGLVKPERDDTEFQHP
                     CFLRGQEQLLENIKRKVTSVSTLKSEDIKIRQDSVTKLLTDVQLMKGKQECMDSKLLA
                     MKHENEALWREVASLRQKHAQQQKVVNKLIQFLISLVQSNRILGVKRKIPLMLNDSGS
                     AHSMPKYSRQFSLEHVHGSGPYSAPSPAYSSSSLYAPDAVASSGPIISDITELAPASP
                     MASPGGSIDERPLSSSPLVRVKEEPPSPPQSPRVEEASPGRPSSVDTLLSPTALIDSI
                     LRESEPAPASVTALTDARGHTDTEGRPPSPPPTSTPEKCLSVACLDKNELSDHLDAMD
                     SNLDNLQTMLSSHGFSVDTSALLDLFSPSVTVPDMSLPDLDSSLASIQELLSPQEPPR
                     PPEAENSSPDSGKQLVHYTAQPLFLLDPGSVDTGSNDLPVLFELGEGSYFSEGDGFAE
                     DPTISLLTGSEPPKAKDPTVS"
BASE COUNT      435 a    739 c    628 g    354 t
ORIGIN


alignment_scores:
            Quality: 2729.00            Length:    529
              Ratio:  5.159               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-304-121-2 x HUMHSF1   ..
```

```
Align seg 1/1  to: HUMHSF1  from: 1  to: 2156

      1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
    161 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 210

     17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
    211 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 260

     34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
    261 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 310

     51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
    311 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

     67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
    361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 410

     84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
    411 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 460

    101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
    461 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 510

    117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
    511 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 560

    134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
    561 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 610

    151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
    611 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

    167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
    661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 710

    184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
    711 AAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAAC 760

    201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
    761 CGGATCCTGGGGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTGG 810

    217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
    811 CTCAGCACATTCCATGCCCAAGTATAGCCGGCAGTTCTCCCTGGAGCACG 860

    234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
    861 TCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTCC 910

    251 SerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAs 267
    911 AGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

    267 pIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerI 284
    961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCA 1010

    284 leAspGluArgProLeuSerSerSerProLeuValArgValLysGluGlu 300
   1011 TAGACGAGAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAG 1060

    301 ProProSerProProGlnSerProArgValGluGluAlaSerProGlyAr 317
   1061 CCCCCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGCG 1110

    317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
   1111 CCCATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCCA 1160

    334 leLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAsp 350
   1161 TCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGAC 1210

    351 AlaArgGlyHisThrAspThrGluGlyArgProProSerProProProTh 367
   1211 GCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

    367 rSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGluL 384
   1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGC 1310

    384 euSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThr 400
   1311 TCAGTGACCACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACC 1360

    401 MetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLe 417
   1361 ATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCT 1410

    417 uPheSerProSerValThrValProAspMetSerLeuProAspLeuAspS 434
   1411 GTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACA 1460

    434 erSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProArg 450
   1461 GCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGG 1510

    451 ProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValHi 467
   1511 CCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

    467 sTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThrG 484
   1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCG 1610

    484 lySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPhe 500
   1611 GGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTC 1660

    501 SerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGl 517
   1661 TCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGG 1710

    517 ySerGluProProLysAlaLysAspProThrValSer 529
   1711 CTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 1747

seq_name: gb_ro:MMHSF1

seq_documentation_block:
LOCUS       MMHSF1       1947 bp    mRNA            ROD       28-NOV-1991
DEFINITION  M.musculus mRNA for heat shock transcription factor 1.
ACCESSION   X61753
VERSION     X61753.1  GI:51445
KEYWORDS    heat shock transcription factor; HSF1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1947)
  AUTHORS   Sarge,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1991) K.D. Sarge, Northwestern University, Dep of
```

```
           Biochem.,Mol. and Cell. Biology, Hogan 2-100, 2153 Sheridan Rd.,
           Evanston IL 60208, USA
REFERENCE  2  (bases 1 to 1947)
  AUTHORS  Sarge,K.D., Zimarino,V., Holm,K., Wu,C. and Morimoto,R.I.
  TITLE    Cloning and characterization of two mouse heat shock factors with
           distinct inducible and constitutive DNA-binding ability
  JOURNAL  Genes Dev. 5 (10), 1902-1911 (1991)
  MEDLINE  92009180
COMMENT    See also x61754.
FEATURES             Location/Qualifiers
     source          1. .1947
                     /organism="Mus musculus"
                     /strain="WEHI-3"
                     /db_xref="taxon:10090"
                     /dev_stage="adult"
                     /cell_type="macrophage precursor"
                     /cell_line="WEHI-3"
                     /clone_lib="lambda ZAP cDNA"
                     /clone="C12"
     CDS             140. .1651
                     /codon_start=1
                     /product="heat shock transcription factor 1"
                     /protein_id="CAA43892.1"
                     /db_xref="GI:51446"
                     /db_xref="MGD:MGI:96238"
                     /db_xref="SWISS-PROT:P38532"
                     /translation="MDLAVGPGAAGPSNVPAFLTKLWTLVSDPDTDALICWSPSGNSF
                     HVFDQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVHIEQGGLVKPERDDTEFQHP
                     CFLRGQEQLLENIKRKVTSVSTLKSEDIKIRQDSVTRLLTDVQLMKGKQECMDSKLLA
                     MKHENEALWREVASLRQKHAQQQKVVNKLIQFLISLVQSNRILGVKRKIPLMLSDSNS
                     AHSVPKYGRQYSLEHVHGPGPYSAPSPAYSSSSLYSSDAVTSSGPIISDITELAPTSP
                     LASPGRSIDERPLSSSTLVRVKQEPPSPPHSPRVLEASPGRPSSMDTPLSPTAFIDSI
                     LRESEPTPAASNTAPMDTTGAQAPALPTPSTPEKCLSVACLDKNELSDHLDAMDSNLD
                     NLQTMLTSHGFSVDTSALLDIQELLSPQEPPRPIEAENSNPDSGKQLVHYTAQPLFLL
                     DPDAVDTGSSELPVLFELGESSYFSEGDDYTDDPTISLLTGTEPHKAKDPTVS"
BASE COUNT      435 a    613 c    521 g    378 t
ORIGIN


alignment_scores:
            Quality: 2294.00            Length:    530
              Ratio:  4.749              Gaps:      3
 Percent Similarity:  91.132   Percent Identity: 85.660

alignment_block:
 US-09-304-121-2 x MMHSF1   ..

 Align seg 1/1  to: MMHSF1  from: 1  to: 1947

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         |||||||||   ||||||||||||||||||||||||||||||||||||||||
     140 ATGGATCTGGCCGTGGGCCCCGGTGCAGCGGGGCCCAGCAACGTCCCGGC 189

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     190 CTTCCTAACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACAGACGCGC 239

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     240 TCATCTGCTGGAGCCCGAGTGGGAACAGCTTCCACGTGTTTGACCAGGGC 289

      51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     290 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 339

      67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     340 TAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGAAAAGTAGTCCACA 389

      84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     390 TTGAGCAGGGTGGCCTGGTCAAGCCTGAGAGAGATGACACCGAGTTCCAG 439

     101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     440 CATCCTTGTTTCTTGCGTGGACAGGAACAGCTCCTTGAGAACATCAAGAG 489

     117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     490 GAAAGTGACCAGCGTGTCCACCCTGAAGAGTGAGGACATAAAAATACGCC 539

     134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
         |||||||||||||:::|||||||||||||||||||||||||||||||||
     540 AGGACAGTGTCACCCGGCTGTTGACAGATGTGCAGCTGATGAAGGGGAAA 589

     151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     590 CAGGAGTGTATGGACTCCAAGCTCCTGGCCATGAAGCACGAGAACGAGGC 639

     167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     640 CCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAGCAAA 689

     184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     690 AAGTTGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCGAAC 739

     201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
         ||||||||||||||||||||||||||||||||||:::|||||||::
     740 CGGATCCTGGGGGTGAAGAGAAAGATCCCTCTGATGTTGAGTGACAGCAA 789

     217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
         :|||||||||||:::|||||||||:::|||||||:::|||||||||||||
     790 CTCAGCACACTCTGTGCCCAAGTATGGTCGACAGTACTCCCTGGAGCATG 839

     234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
         ||||||||   ||||||||||||||||||||||||||||||||||||||||
     840 TCCATGGTCCTGGCCCATACTCAGCTCCATCTCCAGCCTACAGCAGCTCT 889

     251 SerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAs 267
         |||||||||:::   |||||||||:::|||||||||||||||||||||||
     890 AGCCTTTACTCCTCTGATGCTGTCACCAGCTCTGGACCCATAATCTCCGA 939

     267 pIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerI 284
         ||||||||||||||||||:::|||||:::|||||||||||||   ||||
     940 TATCACTGAGCTGGCTCCCACCAGCCCTTTGGCCTCCCCAGGCAGGAGCA 989

     284 leAspGluArgProLeuSerSerSerProLeuValArgValLysGluGlu 300
         |||||||||||||||||||||||||   |||||||||||||||:::|||
     990 TAGATGAGAGGCCTCTGTCCAGCAGCACTCTGGTCCGTGTCAAGCAAGAG 1039

     301 ProProSerProProGlnSerProArgValGluGluAlaSerProGlyAr 317
         |||||||||||||||:::|||||||||||||   |||||||||||||||||
    1040 CCCCCCAGCCCACCTCACAGCCCTCGGGTACTGGAGGCGAGCCCTGGGCG 1089

     317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
         ||||||||||:::||||||   |||||||||||||||:::|||||||||||
    1090 CCCATCCTCCATGGATACCCCTTTGTCCCCAACTGCCTTCATTGACTCCA 1139

     334 leLeuArgGluSerGluProAlaPro...AlaSerValThrAlaLeuThr 349
         ||||||||||||||||||||:::|||   ||||||   ||||||
    1140 TCCTTCGAGAGAGCGAGCCTACCCCTGCTGCCTCAAACACAGCCCCTATG 1189

     350 AspAlaArgGlyHisThrAspThrGluGlyArgProProSerProProPr 366
         |||:::   |||               ::::::   |||:::   |||
    1190 GACACAACCGGA.............GCCCAAGCCCCCGCACTCCCGAC 1224

     366 oThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnG 383
             ||||||||||||||||||||||||||||||||||||||||||||||||
    1225 CCCCTCCACCCCTGAGAAGTGCCTCAGCGTAGCCTGCCTAGACAAGAACG 1274

     383 luLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGln 399
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1275 AGCTAAGTGATCACCTGGATGCCATGGACTCCAACCTGGACAACCTGCAG 1324

     400 ThrMetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAs 416
```

```
     |||||||||:::|||||||||||||||||||||||||||||||||||||||
  1325 ACCATGCTGACAAGCCACGGCTTCAGTGTGGACACCAGTGCCCTGCTGGA 1374

   416 pLeuPheSerProSerValThrValProAspMetSerLeuProAspLeuA 433
       |
  1375 C.............................................. 1375

   433 spSerSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProPro 449
                     ||||||||||||||||||||||||||||||||||||
  1376 ................ATTCAGGAGCTTCTGTCTCCACAAGAGCCTCCC 1408

   450 ArgProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuVa 466
       ||||||   ||||||||||||||||:::|||||||||||||||||||||||
  1409 AGGCCTATTGAGGCAGAGAACAGTAACCCCGACTCAGGAAAGCAGCTGGT 1458

   466 lHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspT 483
       |||||||||||||||||||||||||||||||||||   :::|||||||
  1459 GCACTACACGGCTCAGCCTCTGTTCCTGCTGGATCCTGATGCTGTGGACA 1508

   483 hrGlySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyr 499
       ||||||||:::::|||||||||||||||||||||||||||:::||||||
  1509 CAGGGAGCAGTGAGCTGCCTGTGCTCTTTGAGCTGGGGGAGAGCTCCTAC 1558

   500 PheSerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuTh 516
       ||||||||||||||   ::::::::::||||||||||||||||||||||
  1559 TTCTCTGAGGGGGATGACTACACGGATGATCCCACCATCTCTCTTCTGAC 1608

   516 rGlySerGluProProLysAlaLysAspProThrValSer 529
       ||||:::||||||   |||||||||||||||||||||||
  1609 AGGCACTGAACCCCATAAAGCCAAGGACCCCACTGTCTCC 1648
```

```
seq_name: gb_ro:RRHSF1

seq_documentation_block:
LOCUS       RRHSF1       1647 bp    mRNA            ROD       30-MAR-1995
DEFINITION  R.rattus mRNA for heat shock transcription factor 1.
ACCESSION   X83094
VERSION     X83094.1  GI:758270
KEYWORDS    heat shock transcription factor 1; HSF1 gene.
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 1647)
  AUTHORS   Swamynathan,S.K., Revathi,C.J. and Srinivas,U.K.
  TITLE     Cloning and characterization of rat heat shock transcription factor
            1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1647)
  AUTHORS   Srinivas,U.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1994) U.K. Srinivas, Centre for Cellular and
            Molecular Biology, Habsiguda, Hyderabad, PIN-500 007, INDIA
FEATURES             Location/Qualifiers
     source          1. .1647
                     /organism="Rattus rattus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10117"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /cell_type="hepatocytes"
                     /clone_lib="Stratagene cDNA library"
     gene            1. .1348
                     /gene="HSF1"
     CDS             <1. .1348
                     /gene="HSF1"
                     /codon_start=2
                     /product="heat shock transcription factor 1"
                     /protein_id="CAA58149.1"
                     /db_xref="GI:758271"
                     /db_xref="SPTREMBL:Q63717"
                     /translation="FRKVVHIEQGGLVKPERDDTEFQHPCFLRGQEQLLENIKRKVTS
                     VSTLKSEDIKIRQDSVTRLLTDVQLMKGKQECMDSKLLAMKHENEALWREVASLRQKH
                     AQQQKVVNKLIQFLISLVQSNRILGVKRKIPLMLSDSSSAHSVPKYGRQYSLEHVHGP
                     GPYSAPSPAYSSSSLYSSDAVTSSGPIISDITELAPTSPLASPGRSIDERPLSSSTLV
                     RVKEEPPSPPHSPRVLEASPGRPASMDTPLSPTAFIDSILRESEPTPAASNTAPMDTT
                     GAQAPAHPAPSTPEKCLSVACLDKNELSDHLDAMDSNLDNLQTMLTSHGFSVDTSALL
                     DLFSPSVTMPDMSLPDLDSSLASIQELLSPQEPPRPIEAENSNADSGKQLVHYTAQPL
                     FLLDPDAVDTGSSELPVLFELGESSYFSEGDDYTDDPTISLLTGTEPHKAKDPTVS"
     3'UTR           1349. .1647
BASE COUNT      384 a    512 c    425 g    326 t
ORIGIN


alignment_scores:
           Quality: 1999.00            Length:    453
             Ratio:  4.671               Gaps:      2
 Percent Similarity:  94.481   Percent Identity: 87.859

alignment_block:
 US-09-304-121-2 x RRHSF1   ..

 Align seg 1/1  to: RRHSF1  from: 1  to: 1647

      78 PheArgLysValValHisIleGluGlnGlyGlyLeuValLysProGluAr 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 TTCCGAAAAGTAGTCCACATTGAGCAGGGTGGCCTGGTCAAGCCTGAGAG 51

      94 gAspAspThrGluPheGlnHisProCysPheLeuArgGlyGlnGluGlnL 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 GGATGACACCGAGTTCCAGCATCCATGTTTCCTGCGTGGCCAGGAACAGC 101

     111 euLeuGluAsnIleLysArgLysValThrSerValSerThrLeuLysSer 127
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TCCTTGAGAATATCAAGAGGAAAGTGACCAGCGTGTCCACCCTGAAGAGT 151

     128 GluAspIleLysIleArgGlnAspSerValThrLysLeuLeuThrAspVa 144
         ||||||||||||||||||||||||||||||||:::|||||||||||||||
     152 GAGGACATAAAAATACGCCAGGACAGTGTCACCAGGCTGTTGACAGATGT 201

     144 lGlnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeuLeuAlaM 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 GCAGCTGATGAAGGGGAAACAGGAGTGTATGGACTCCAAGCTCCTGGCCA 251

     161 etLysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGln 177
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TGAAGCACGAGAACGAGGCCCTGTGGCGGGAGGTGGCCAGCCTTCGGCAG 301

     178 LysHisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIl 194
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 AAGCATGCCCAGCAGCAAAAAGTTGTCAACAAGCTCATCCAATTCCTGAT 351

     194 eSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleProL 211
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 CTCACTGGTGCAGTCGAACCGGATCCTGGGGGTGAAAAGAAAGATCCCTC 401

     211 euMetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSerArg 227
         ||||||||:::|||||:::|||||||||||||:::|||||||||:::|||
     402 TGATGTTGAGTGACAGCAGCTCAGCACACTCTGTGCCCAAGTATGGTCGA 451

     228 GlnPheSerLeuGluHisValHisGlySerGlyProTyrSerAlaProSe 244
         |||:::|||||||||||||||||||||   ||||||||||||||||||||
     452 CAGTACTCCCTGGAGCATGTCCATGGTCCCGGCCCATACTCAGCTCCATC 501

     244 rProAlaTyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSerS 261
         ||||||||||||||||||||||||||:::   |||||||||:::||||||
     502 TCCAGCCTACAGCAGCTCTAGCCTTTACTCCTCTGATGCTGTCACCAGCT 551

     261 erGlyProIleIleSerAspIleThrGluLeuAlaProAlaSerProMet 277
         ||||||||||||||||||||||||||||||||||:::|||||||:::
     552 CTGGGCCCATAATCTCCGATATCACTGAGCTGGCTCCCACCAGCCCTTTG 601

     278 AlaSerProGlyGlySerIleAspGluArgProLeuSerSerSerProLe 294
         ||||||||||||   |||||||||||||||||||||||||||||   ||
     602 GCCTCCCCAGGCAGGAGCATAGATGAGAGGCCTCTGTCCAGCAGCACCCT 651
```

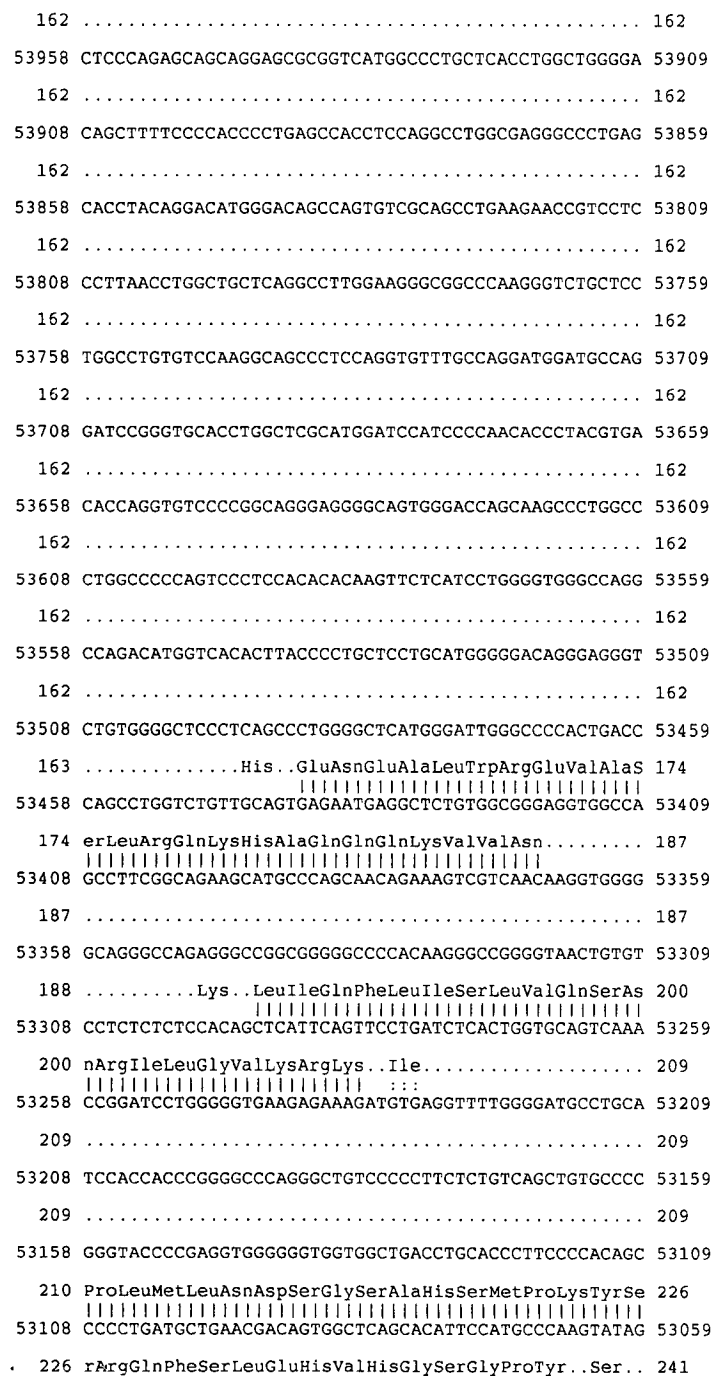
```
   294 uValArgValLysGluGluProProSerProProGlnSerProArgValG 311
       ||||||||||||||||||||||||||||||||||:::|||||||||||||
   652 GGTCCGTGTCAAGGAAGAGCCCCCCAGCCCACCTCATAGCCCTCGGGTAC 701

   311 luGluAlaSerProGlyArgProSerSerValAspThrLeuLeuSerPro 327
       ||||||||||||||||||||||:::|||:::||||||   |||||||||
   702 TGGAGGCAAGCCCTGGGCGCCCAGCCTCCATGGATACCCCTTTGTCCCCA 751

   328 ThrAlaLeuIleAspSerIleLeuArgGluSerGluProAlaPro...Al 343
       ||||||:::||||||||||||||||||||||||||||:::|||   ||
   752 ACTGCCTTCATTGACTCCATCCTTCGAGAGAGCGAACCTACCCCTGCTGC 801

   343 aSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyA 360
       ||||   ||||||      |||:::   |||             ::::
   802 CTCAAACACAGCCCCTATGGACACAACCGGA.............GCCC 836

   360 rgProProSerProProProThrSerThrProGluLysCysLeuSerVal 376
       ::   |||:::   |||      ||||||||||||||||||||||||||
   837 AAGCCCCTGCACACCCAGCCCCCTCCACCCCTGAGAAGTGCCTCAGCGTA 886

   377 AlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSe 393
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   887 GCCTGCCTAGACAAGAACGAGCTAAGTGATCACCTGGATGCCATGGACTC 936

   393 rAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValA 410
       ||||||||||||||||||||||||||||:::||||||||||||||||||
   937 CAACCTGGACAACCTGCAGACCATGCTGACAAGCCACGGCTTCAGTGTGG 986

   410 spThrSerAlaLeuLeuAspLeuPheSerProSerValThrValProAsp 426
       ||||||||||||||||||||||||||||||||||||:::||||||
   987 ACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCATGCCCGAC 1036

   427 MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLe 443
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1037 ATGAGCCTGCCTGACCTGGACAGCAGCCTGGCCAGCATTCAGGAACTTCT 1086

   443 uSerProGlnGluProProArgProProGluAlaGluAsnSerSerProA 460
       ||||||||||||||||||||||||   ||||||||||||||:::   |
  1087 ATCTCCCCAAGAGCCTCCCAGGCCTATTGAGGCAGAGAATAGTAACGCCG 1136

   460 spSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeu 476
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1137 ACTCAGGAAAGCAGCTGGTGCACTACACGGCTCAGCCTCTGTTCCTGCTG 1186

   477 AspProGlySerValAspThrGlySerAsnAspLeuProValLeuPheGl 493
       ||||||   :::||||||||||||||:::::::||||||||||||||||||
  1187 GACCCTGATGCTGTGGACACAGGGAGCAGTGAACTGCCTGTGCTCTTTGA 1236

   493 uLeuGlyGluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAspP 510
       ||||||||||:::|||||||||||||||||||||   :::::::::||||
  1237 GCTGGGGGAGAGCTCCTACTTCTCTGAGGGAGATGACTACACGGATGATC 1286

   510 roThrIleSerLeuLeuThrGlySerGluProProLysAlaLysAspPro 526
       ||||||||||||||||||||||:::|||||||   ||||||||||||||||
  1287 CCACCATCTCTCTTCTGACGGGCACTGAACCCCACAAAGCCAAGGACCCC 1336

   527 ThrValSer 529
       |||||||||
  1337 ACTGTCTCC 1345

seq_name: gb_ov:CHKHSF3A

seq_documentation_block:
LOCUS       CHKHSF3A     1774 bp    mRNA            VRT       17-JUN-1993
DEFINITION  Chicken mRNA sequence.
ACCESSION   L06098
VERSION     L06098.1  GI:289814
KEYWORDS    .
SOURCE      Gallus gallus (sub_species domesticus) (library: lambda ZAPII)
            embryo blood cDNA to mRNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE   1  (bases 1 to 1774)
  AUTHORS   Nakai,A. and Morimoto,R.I.
  TITLE     Characterization of a novel chicken heat sock transcription factor,
            heat shock factor 3, suggests a new regulatory pathway
  JOURNAL   Mol. Cell. Biol. 13, 1983-1997 (1993)
  MEDLINE   93204945
FEATURES             Location/Qualifiers
     source          1. .1774
                     /organism="Gallus gallus"
                     /sub_species="domesticus"
                     /db_xref="taxon:9031"
                     /cell_type="red blood cell"
                     /dev_stage="embryo"
                     /tissue_type="blood"
                     /tissue_lib="lambda ZAPII"
BASE COUNT      443 a    582 c    479 g    270 t
ORIGIN

alignment_scores:
            Quality: 1982.50            Length:    532
              Ratio:  4.386              Gaps:      5
 Percent Similarity:  84.962   Percent Identity: 72.932

alignment_block:
 US-09-304-121-2 x CHKHSF3A   ..

 Align seg 1/1  to: CHKHSF3A  from: 1  to: 1774

      6 GlyProGlyAlaAla.....................GlyProSerAs 14
        |||||||||||||||                      |||   |||||
     37 GGCCCCGGCGCGGCGGCGGCGGCTGTGGGAGCGGGCCCGGGCGGCAGCAA 86

     14 nValProAlaPheLeuThrLysLeuTrpThrLeuValSerAspProAspT 31
        ||||   |||||||||||||||||||||||||||:::||||||:::|
     87 CGTGTCGGCCTTCCTGACCAAGCTGTGGACGTTGGTGGAGGACCCCGAGA 136

     31 hrAspAlaLeuIleCysTrpSerProSerGlyAsnSerPheHisValPhe 47
        |||||   ||||||||||||||||||||||||||||||||||||||||||
    137 CGGACCCGCTGATCTGCTGGAGCCCGAGTGGAAACAGCTTCCATGTGTTC 186

     48 AspGlnGlyGlnPheAlaLysGluValLeuProLysTyrPheLysHisAs 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    187 GACCAAGGACAGTTCGCCAAGGAGGTGCTCCCAAAGTACTTCAAGCACAA 236

     64 nAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysV 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    237 CAACATGGCCAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAGG 286

     81 alValHisIleGluGlnGlyGlyLeuValLysProGluArgAspAspThr 97
        ||||||||||||||||||||||||||||||||||||||||||:::||||||||||
    287 TGGTGCACATCGAGCAGGGAGGGTTGGTGAAGCCAGAGAAGGACGACACT 336

     98 GluPheGlnHisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAs 114
        ||||||||||||||   |||:::|||||||||||||:::|||||||||||
    337 GAGTTCCAGCACCCCTATTTCATCCGTGGCCAGGAGCACCTCCTGGAGAA 386

    114 nIleLysArgLysValThrSerValSerThrLeuLysSerGluAspIleL 131
        ||||||||||||||||||||||||||||||:::::|||:::||||||||||
    387 CATCAAGAGGAAAGTCACCAGCGTGTCCAGCATAAAGAACGAGGACATCA 436

    131 ysIleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMet 147
        ||:::|||||||||:::|||||||||||||||||||:::|||:::|||
    437 AGGTGAGGCAAGACAACGTCACCAAGCTGCTGACCGACATCCAGGTGATG 486

    148 LysGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGl 164
        ||||||||||||||   ||||||||||||||:::|||||||||||||||
    487 AAAGGAAAGCAGGAGAGCATGGACTCCAAGCTGATAGCCATGAAGCATGA 536

    164 uAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaG 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    537 GAACGAGGCGCTGTGGAGGGAGGTGGCCAGCCTAAGGCAGAAGCACGCCC 586
```

```
 181 lnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleSerLeuVal 197
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 587 AGCAGCAGAAGGTGGTCAATAAGCTCATCCAGTTCTTGATTTCATTGGTG 636

 198 GlnSerAsnArgIleLeuGlyValLysArgLysIleProLeuMetLeuAs 214
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 637 CAATCCAACCGCATCCTCGGGGTGAAGAGGAAGATCCCCCTGATGTTGAA 686

 214 nAspSerGlySerAlaHisSerMetProLysTyrSerArgGlnPheSerL 231
     |||||||:::|||||||||||||||||||||||||||||||:::||||
 687 CGACAGCAGTTCAGCTCATTCCATGCCGAAATACAGCCGCCAGTACTCCC 736

 231 euGluHisValHisGlySerGlyProTyrSerAlaProSerProAlaTyr 247
     ||||||||||||||||||:::|||||:::|||   ||||||||||||
 737 TGGAGCACGTCCACGGCTCATCCCCATACGCAGCTTCTTCCCCAGCGTAC 786

 248 SerSerSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIl 264
     |||:::|||:::::::|||:::||||||:::   ::::::||||||||||
 787 AGCGGCTCCAACATCTACTCCCCCGATTCATCCACCAACTCCGGCCCCAT 836

 264 eIleSerAspIleThrGluLeuAlaProAlaSerProMetAlaSerProG 281
     ||||||||||:::||||||||||||   :::||||||   |||||||||:
 837 CATCTCGGACGTGACGGAGCTGGCCCAGTCCAGCCCTTCAGCATCTCCCA 886

 281 lyGlySerIleAspGluArgProLeuSerSerSerProLeuValArgVal 297
     ::|||||:::|||||||||          |||||||||:::||||||:::
 887 GTGGCAGCCTGGATGAGAGG.........AGCAGCCCTGTGGTCCGCATC 927

 298 LysGluGluProProSerProProGlnSerProArgValGluGluAlaSe 314
     ||||||||||||||||||   :::|||||||
 928 AAAGAGGAGCCCCCCAGCCCCTCCCGCAGCCCC..................  960

 314 rProGlyArgProSerSerValAspThrLeuLeuSerProThrAlaLeuI 331

 960 .................................................. 960

 331 leAspSerIleLeuArgGluSerGluProAlaProAlaSerValThrAla 347
                 :::|||:::|||||               ::::::
 961 ..............AAGGAGAACGAGCCC..............AGCACC 981

 348 LeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProProSerPr 364
        |||   |||   |||:::::      :::   :::|||
 982 ACGACGGCCGCGGCCGGCAACAGCACCGAGCAGCCTCAGCCC........ 1023

 364 oProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspL 381
                     ||||||||||||||||||||||||||||||||
1024 ..............CAGGAAAAGTGCCTCAGCGTTGCCTGCCTCGACA 1057

 381 ysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsn 397
     ||||||||||:::|||||||||||:::::|||||||||||||||||||
1058 AGAACGAACTGAACGACCACCTGGACACCATCGACTCCAACCTGGACAAC 1107

 398 LeuGlnThrMetLeuSerSerHisGlyPheSerValAspThrSerAlaLe 414
     ||||||||||||||||:::|||||||||||||||||||||:::|||||
1108 CTGCAGACCATGCTGAGCACGCACGGCTTCAGCGTCGACACCACCGCCCT 1157

 414 uLeuAspLeuPheSerProSerValThrValProAspMetSerLeuProA 431
     ||||||||||||||||||||||:::||||||   ||||||:::|||||||
1158 GCTGGACCTCTTCAGCCCCTCCATGACGGTTACGGATATGAACCTCCCCG 1207

 431 spLeuAspSerSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGlu 447
     ||||||||||||||||||||||||||:::|||||||||   ||||||
1208 ACCTGGACAGCAGCCTCGCCAGCATCCAGGATCTGCTCTCGTCGCAGGAG 1257

 448 ProProArgProProGluAlaGluAsnSerSerProAspSerGlyLysGl 464
         :::|||   ||||||:::   ::::::   |||:::|||||||
1258 CAGCAGAAGCCCTCGGAGGCCGATGCGGCAGCAGCAGACACAGGGAAGCA 1307

 464 nLeuValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerV 481
     ||||||||||||||||||||||||||||||||:::|||   ::::::|
1308 GTTGGTTCACTACACGGCTCAGCCCCTCTTCCTGGTGGATTCGAGCGCGG 1357

 481 alAspThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGluGly 497
```

```
     |||||:::||||||:::|||||||||:::::|||||||||||||||||||
1358 TGGACGTGGGCAGCGGAGACCTCCCCATCTTCTTCGAGTTGGGGGAGGGC 1407

 498 SerTyrPheSerGluGlyAspGlyPheAlaGluAspProThrIleSerLe 514
     |||||||||:::::|||||   :::   ||||||||||||||||||||
1408 TCCTATTTCACCGACGGGGACGAATACAACGAAGACCCCACCATCTCCCT 1457

 514 uLeuThrGlySerGluProProLysAlaLysAspProThrValSer 529
     ||||:::|||:::|||   ||||||   ||||||||||||||||||
1458 CCTATCCGGGACTGAGCAACCCAAACCCAAGGACCCCACCGTGTCC 1503

seq_name: gb_htg6:AF205589

seq_documentation_block:
LOCUS       AF205589   131973 bp    DNA             HTG       07-DEC-1999
DEFINITION  Homo sapiens chromosome 8 clone BAC 393G12, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AF205589
VERSION     AF205589.1  GI:6531668
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 131973)
  AUTHORS   Polley,A., Wen,G., Baumgart,C., Dette,M., Jahn,N., Schilhabel,M.,
            Menzel,U. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     1-16995: contig of 16995 bp;  16995-16996: gap ofunknown size;
            16996-28004: contig of 11009 bp;  28004-28005: gap of unknownsize;
            28005-43979: contig of 15975 bp;  43979-43980: gap of unknown size;
             43980-67713: contig of 23734 bp;  67713-67714: gap of unknown
            size;  67714-77798: contig of 10085 bp;  77798-77799: gap of
            unknown size;  77799-92762: contig of 14964 bp;  92762-92763: gap
            of unknown size;  92763-98839: contig of 6077 bp;  98839-98840: gap
            of unknown size;  98840-104153: contig of 5314 bp;  104153-104154:
            gap of unknown size;  104154-108012: contig of 3859 bp;
            108012-108013: gap of unknown size;  108013-111572: contig of 3560
            bp;  111572-111573: gap of unknown size;  111573-113894: contig of
            2322 bp;  113894-113895: gap of unknown size;  113895-131973:
            contig of 18079 bp;.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .131973
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="BAC 393G12"
                     /note="seed clone"
     exon            complement(213. .614)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            240. .541
                     /note="MZEF"
                     /evidence=not_experimental
     exon            240. .541
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            240. .541
                     /note="GENSCAN"
     exon            240. .547
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            760. .931
                     /note="MZEF"
                     /evidence=not_experimental
     exon            760. .931
                     /note="GENSCAN"
```

```
     exon            760. .871
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            763. .891
                     /note="XPOUND"
                     /evidence=not_experimental
     repeat_region   1011. .1136
                     /rpt_family="FLAM_A"
     misc_feature    1138. .1441
                     /note="CpGisland: CpG > 5%"
     repeat_region   1144. .1439
                     /rpt_family="AluY"
     exon            1574. .1759
                     /note="GENSCAN"
     exon            2040. .2234
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            2059. .2234
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            2059. .2234
                     /note="GENSCAN"
     exon            2182. .2234
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   2394. .2701
                     /rpt_family="AluSg"
     repeat_region   2707. .3012
                     /rpt_family="AluY"
     exon            3368. .3441
                     /note="GENSCAN"
     exon            3982. .4115
                     /note="GENSCAN"
     exon            4019. .4115
                     /note="XPOUND"
                     /evidence=not_experimental
     misc_feature    5092. .6790
                     /note="CpGisland: CpG > 5%"
     exon            5107. .5159
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            5162. .5376
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            5183. .5376
                     /note="GENSCAN"
     exon            5183. .5376
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   5431. .5735
                     /rpt_family="AluY"
     misc_feature    5774. .6055
                     /note="CpGisland: CpG > 10%"
     exon            5955. .6030
                     /note="GENSCAN"
     exon            complement(6254. .6367)
                     /note="XPOUND"
                     /evidence=not_experimental
     repeat_region   6573. .6730
                     /rpt_family="L2"
     repeat_region   complement(7193. .7499)
                     /rpt_family="AluSx"
     exon            7920. .7955
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            7920. .7955
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            8311. .8339
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8311. .8339
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8311. .8339
                     /note="GENSCAN"
     exon            8369. .8493
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            8404. .8493
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8456. .8493
                     /note="GENSCAN"
     exon            8456. .8493
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8731. .8836
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8738. .8836
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            8747. .8836
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8747. .8836
                     /note="GENSCAN"
     exon            9468. .9573
                     /note="MZEF"
                     /evidence=not_experimental
     exon            9468. .9573
                     /note="GENSCAN"
     exon            9479. .9573
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            9479. .9573
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            9700. .9801
                     /note="MZEF"
                     /evidence=not_experimental
     exon            9700. .9801
                     /note="GENSCAN"
     exon            9700. .9801
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            9700. .9801
                     /note="GRAIL"
                     /evidence=not_experimental
     misc_feature    10188. .10698
                     /note="CpGisland: CpG > 5%"
     exon            10203. .10304
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            complement(10241. .10371)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10251. .10304
                     /note="GENSCAN"
     exon            10251. .10334
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10251. .10304
                     /note="MZEF"
                     /evidence=not_experimental
     exon            10345. .10478
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10387. .10475
                     /note="XPOUND"

alignment_scores:
           Quality: 1538.50                Length:   1803
             Ratio:  3.179                   Gaps:     19
 Percent Similarity:  26.844   Percent Identity: 26.567
```

```
alignment_block:
 US-09-304-121-2 x AF205589/rev   ..

 Align seg 1/1  to reverse of: AF205589  from: 1  to: 131973

     40 SerGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGluVa 56
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  55708 AGCGGGAACAGCTTCCACGTGTTCGACCAGGGCCAGTTTGCCAAGGAGGT 55659

     56 lLeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGlnL 73
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  55658 GCTGCCCAAGTACTTCAAGCACAACAACATGGCCAGCTTCGTGCGGCAGC 55609

     73 euAsnMet..................................... 75
        ||||||||
  55608 TCAACATGTGTGAGTGCTGCCTCCAGGCAGCGCAGGGGTGCGGGAGGCAG 55559

     75 ..................................................... 75

  55558 ACCTGCAGATGGCGGGACCCCAGCAGGAGGACCCTGTGATGAAGGACGGG 55509

     75 ..................................................... 75

  55508 GCGTCCTGTGCTGGCCAAGGGCATGGCGTGGGACCCTGCAAGCCTGGGGG 55459

     75 ..................................................... 75

  55458 CCGGGGAGCAGCCGCCTCTTCCATGGGGGAGGGTCCTTGTGGGTATGAAC 55409

     75 ..................................................... 75

  55408 CTGGGGTCCCCATGGAAGAACCGTGAAGCCGGAGCTGTACTCCACGTGTG 55359

     75 ..................................................... 75

  55358 TCGGGCGCAGGGAGCCCTGTGGGGACACAGGGTCTCCCTTAGACCAAGGC 55309

     75 ..................................................... 75

  55308 CACTCGGCCACCCAGGCATGGGCTCTGAGGGGGCAGGGCAGGGTCTGACC 55259

     75 ..................................................... 75

  55258 ATGGCCAAGCCCCGCAGCAGCCTCCTGGAGCAGTGGCCGCTCTTCAGGGG 55209

     76 .....................................................Tyr. 76

  55208 TTCTGGTCCCGCCCTGAGGCAGAGCTGCCCCCTTCCCTGTTATGTGCAGA 55159

     77 .GlyPheArgLysValValHisIleGluGlnGlyGlyLeuValLysProG 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  55158 TGGCTTCCGGAAAGTGGTCCACATCGAGCAGGGCGGCCTGGTCAAGCCAG 55109

     93 luArgAspAspThrGluPheGlnHisProCysPheLeuArgGlyGlnGlu 109
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  55108 AGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCAGGAG 55059

    110 GlnLeuLeuGluAsnIleLysArgLysValThrSer.............. 121
        ||||||||||||||||||||||||||||||||||
  55058 CAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGAGTGCCGGCCC 55009

    121 ..................................................... 121

  55008 TGCACCCTTGCCCGGTCTCACCTGCCACAGCTCTCCCCGCCCGCCCCTCC 54959

    121 ..................................................... 121

  54958 CTGAGGGCTCCCACCCCAGCCTGCCCCTTCCTGAGGGACCTGGCTGCAGT 54909

    121 ..................................................... 121

  54908 GGACGAGCCTGAGCCTGCCAAGCTCCTAGGCTGCTCCAGTGACTCCTGTC 54859

    122 ................ValSerThrLeuLysSerGluAspIleLysIleA 133
                         ||||||||||||||||||||||||||||||||||
  54858 CCTCTCGGGAATCCAGGTGTCCACCCTGAAGAGTGAAGACATAAAGATCC 54809

    133 rgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGly 149
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  54808 GCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGG 54759

    150 LysGlnGluCysMetAspSerLysLeuLeuAlaMetLys........... 162
        ||||||||||||||||||||||||||||||||||||
  54758 AAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCAGTAGGTCCC 54709

    162 ..................................................... 162

  54708 ACACCAGCATTATGGGCCACAGCGGGTCCTGGCGCCCACCAAGAGGCCCC 54659

    162 ..................................................... 162

  54658 GGGTGCTGTGGGGGCAGGGCCCTGACCGGGGCCAGGTGCTCAGCTCCACC 54609

    162 ..................................................... 162

  54608 CTCCCGCTCCACGCACATCTACCCTGGGCCTCTGCCCCAGCCCCGCCGCC 54559

    162 ..................................................... 162

  54558 CGTGGCTGCTGCAACAGCTCTGGGGCCAGCCGTTTTCCATGTGCAGAAGG 54509

    162 ..................................................... 162

  54508 GGGTTGGGAGGGTCCCCTGCTCTGCACATCCCCAGCGCCCTCTGGGCCAT 54459

    162 ..................................................... 162

  54458 TGGCTTCTGCCTGGCTCCGAGCTTGGCGGGACCCACCAGCCTACTTTTGA 54409

    162 ..................................................... 162

  54408 GGACAAGCCTTGCACAGCAGACCCGTCCCACCCGCCTCGGGCCTGGGGTC 54359

    162 ..................................................... 162

  54358 AGACCCACCTCAGGCCCTTCCTGGAACCGCTGTCCCACACTCGCCTGGGC 54309

    162 ..................................................... 162

  54308 CCATGGAGGTAGCCCCCAGCCCCTCCTCACCGGGACCCTGCGGTTCTGGG 54259

    162 ..................................................... 162

  54258 GAAACTTTAGCAGGAGCTTAACACACGGTGGCACCTGAGAAACTGAGGCA 54209

    162 ..................................................... 162

  54208 AGCCCACCTCGGAAGGCCAGCTCAGGTCTCACGTCCACTTTGGGGCTCCA 54159

    162 ..................................................... 162

  54158 GGCATGCGGGAACCCCAATGCCAGTACGGCTTCCTCGGCTCCGTGCAGCA 54109

    162 ..................................................... 162

  54108 GGCGTTCTTGACCCCCACGTCATGGAGGAGGCCTTGGCCTCAGACTATCC 54059

    162 ..................................................... 162

  54058 TGCATGGGGTCTAGCAGGAGCCACCGGCTTTGTGATTCCTGGAGTCCGGC 54009

    162 ..................................................... 162

  54008 CACGTGGACCTGGCCTACCAAGAAGACTCACAGGCTTGACAGGGCCCAGC 53959
```


162 162
53958 CTCCCAGAGCAGCAGGAGCGCGGTCATGGCCCTGCTCACCTGGCTGGGGA 53909
162 162
53908 CAGCTTTTCCCCACCCCTGAGCCACCTCCAGGCCTGGCGAGGGCCCTGAG 53859
162 162
53858 CACCTACAGGACATGGGACAGCCAGTGTCGCAGCCTGAAGAACCGTCCTC 53809
162 162
53808 CCTTAACCTGGCTGCTCAGGCCTTGGAAGGGCGGCCCAAGGGTCTGCTCC 53759
162 162
53758 TGGCCTGTGTCCAAGGCAGCCCTCCAGGTGTTTGCCAGGATGGATGCCAG 53709
162 162
53708 GATCCGGGTGCACCTGGCTCGCATGGATCCATCCCCAACACCCTACGTGA 53659
162 162
53658 CACCAGGTGTCCCCGGCAGGGAGGGGCAGTGGGACCAGCAAGCCCTGGCC 53609
162 162
53608 CTGGCCCCCAGTCCCTCCACACACAAGTTCTCATCCTGGGGTGGGCCAGG 53559
162 162
53558 CCAGACATGGTCACACTTACCCCTGCTCCTGCATGGGGGACAGGGAGGGT 53509
162 162
53508 CTGTGGGGCTCCCTCAGCCCTGGGGCTCATGGGATTGGGCCCCACTGACC 53459
163 .....His..GluAsnGluAlaLeuTrpArgGluValAlaS 174
53458 CAGCCTGGTCTGTTGCAGTGAGAATGAGGCTCTGTGGCGGGAGGTGGCCA 53409
174 erLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsn..... 187
53408 GCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAACAAGGTGGGG 53359
187 187
53358 GCAGGGCCAGAGGGCCGGCGGGGGCCCCACAAGGGCCGGGGTAACTGTGT 53309
188 .....Lys..LeuIleGlnPheLeuIleSerLeuValGlnSerAs 200
53308 CCTCTCTCTCCACAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAA 53259
200 nArgIleLeuGlyValLysArgLys..Ile..... 209
53258 CCGGATCCTGGGGGTGAAGAGAAAGATGTGAGGTTTTGGGGATGCCTGCA 53209
209 209
53208 TCCACCACCCGGGGCCCAGGGCTGTCCCCCTTCTCTGTCAGCTGTGCCCC 53159
209 209
53158 GGGTACCCCGAGGTGGGGGGTGGTGGCTGACCTGCACCCTTCCCCACAGC 53109
210 ProLeuMetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSe 226
53108 CCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAG 53059
226 rArgGlnPheSerLeuGluHisValHisGlySerGlyProTyr..Ser.. 241
53058 CCGGCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGTGA 53009
241 241
53008 GTGCCGGAGACAGGGCACCCGCCCAGGCATGCAGGCGGCAGTGGGGTGGG 52959
242 .....AlaPro 243
52958 TTGCCCCTGCCGGCACTGCATGGACCTCCTGCCTTTGATTGCAGGCCCCC 52909
244 SerProAlaTyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSe 260
52908 TCCCCAGCCTACAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAG 52859
260 rSerGlyProIleIleSerAspIleThrGluLeuAlaProAlaSerProM 277
52858 CTCTGGACCCATCATCTCCGACATCACCGAGCTGGCTCCTGCCAGCCCCA 52809
277 etAlaSerProGlyGlySerIleAspGlu..... 286
52808 TGGCCTCCCCCGGCGGGAGCATAGACGAGAGGTGGGGGCCGCATCACCCC 52759
286 286
52758 AGCCATCCTGTCCCCCAATGAGGCGAGCCCCTGTGGGCCCAGGGCAGAGT 52709
286 286
52708 TGGGGATGAGGTGGGGCTGGCCGAGACGCCAGCTCACCTGGCCCCCCTCG 52659
287 .....ArgProLeuSerSerSerProLeuValArgValLysGluGluPro 301
52658 TGTGCAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCC 52609
302 ProSerProProGlnSerProArgValGluGluAlaSerProGlyArgPr 318
52608 CCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCC 52559
318 oSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerIleL 335
52558 ATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCCATCC 52509
335 euArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAspAla 351
52508 TGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGACGCC 52459
352 ArgGlyHisThrAspThrGluGlyArgProProSerProProProThrSe 368
52458 AGGGGCCACACGGACACCGAGGGCCGGCTTCCTTCCCCCCCGCCCACTTC 52409
368 rThrProGluLysCysLeuSerValAlaCysLeuAspLys..... 381
52408 CACCCATGAAAAGTGCCTTAGGGTAGCCTGCCTGGACAAGTGAGTGCCGC 52359
381 381
52358 CCACCCCTGGCCCCACCCACAGCGCCTGGACGCACAGCCCTGGGCTTCAG 52309
381 381
52308 CCCCGACTGTCCCAGTGGACTGAGCAGGGCAGCTGGCGAGGCAGGACCCT 52259
381 381
52258 ACCCCCAACCTCGGAGCTCGGGGCTGGGGAGGGAGACAGGTGCCAGCCAG 52209
381 381
52208 ATCACCCCACAGTCCCAAACGCCACAGAAGCCACGGGCATGTCCGGACAG 52159
381 381

```
52158 GCTGCCGGGCCCTGCTCTCAGCCACCCCTCCTGCTGCCGCCACCCTCGCC 52109
  381 ..................................................... 381
52108 ACAGGCCACGGGCCCTGGCAGGTCGTGCTGCCCAGACACATGGCAGGAGA 52059
  381 ..................................................... 381
52058 TGGTGAGCAGGGCCGGCCTCCACACCCCCAGCCCCTGCCTGCAATGGGGG 52009
  381 .....:............................................... 381
52008 CTCTTGTTTTTGTATCTTGCAGTTTGGCTCGCACTCCACAGATGTCTAGG 51959
  381 ..................................................... 381
51958 GTCGCCCGCCTCTTCCCCTGCCCCTCTTCCTCTCCGCATGGCCAAGTCCA 51909
  381 ..................................................... 381
51908 GCCAGGGTTAGCGCTGACCCCACTGGGGAGGGAGGAGGGCTCTGCCAGCG 51859
  381 ..................................................... 381
51858 CTCGGGCCCTCCCACACAGCCGTGGACCAGACCCAGCCTGGCCGGGCATG 51809
  381 ..................................................... 381
51808 AGCGTCGGGCCTGGGCGGCAGCAGCCGAGACCCCTCTGTGGCGGGGGCTC 51759
  381 ..................................................... 381
51758 AGTGTCGTCATGGCTCCCCTGGCCACGGGCCTGTGGTCATTGCCTGTGAC 51709
  381 ..................................................... 381
51708 AGGGATGACAGGGACAGGCCTCCCTGCTCTGGCACGGCCTCTGGTGTTGG 51659
  381 ..................................................... 381
51658 CAGGATCCAAACTGCGGTGCTGACTGCACTTCCTGTCAGGCAGGGTCTCC 51609
  381 ..................................................... 381
51608 AGGGCACCTCTGGGACCCAGCCTGGCTGTCCGTCCCATAGCCCAGGTGTG 51559
  381 .....:............................................... 381
51558 CCCTGTGGCCTAGGGCTTTCTTGGTCACAGCCACCAGACCGTGGGTCTCA 51509
  381 ..................................................... 381
51508 TCCCCACAGCAGCACCAGGGCAGGGACTGATCCACCCACACACTGAGCAG 51459
  381 ..................................................... 381
51458 AGCCCCACCTGTTTTCCCAGTGCAACGGGAAACCTCACCAACCCTGAACA 51409
  381 ..................................................... 381
51408 CTGAAATCCCTGATCCTTGTGCTGGAGCTGAATACGCCCCTTGTCTCCTG 51359
  381 ..................................................... 381
51358 GGTCCACTGCCACCAAGGCCCCCAGGCCCTCATGGCAAAGGGATGCCCAG 51309
  381 ..................................................... 381
51308 CATAGGCCCAGCCGCACCCCTGCAGGGCACAGAGCCTCCACTGCCTTCCA 51259
  381 ..................................................... 381
51258 GGCCGTCCTCGAATGCGCTGCCCCCTCCAGCCTGTGCAGGCGTACACGGG 51209
```

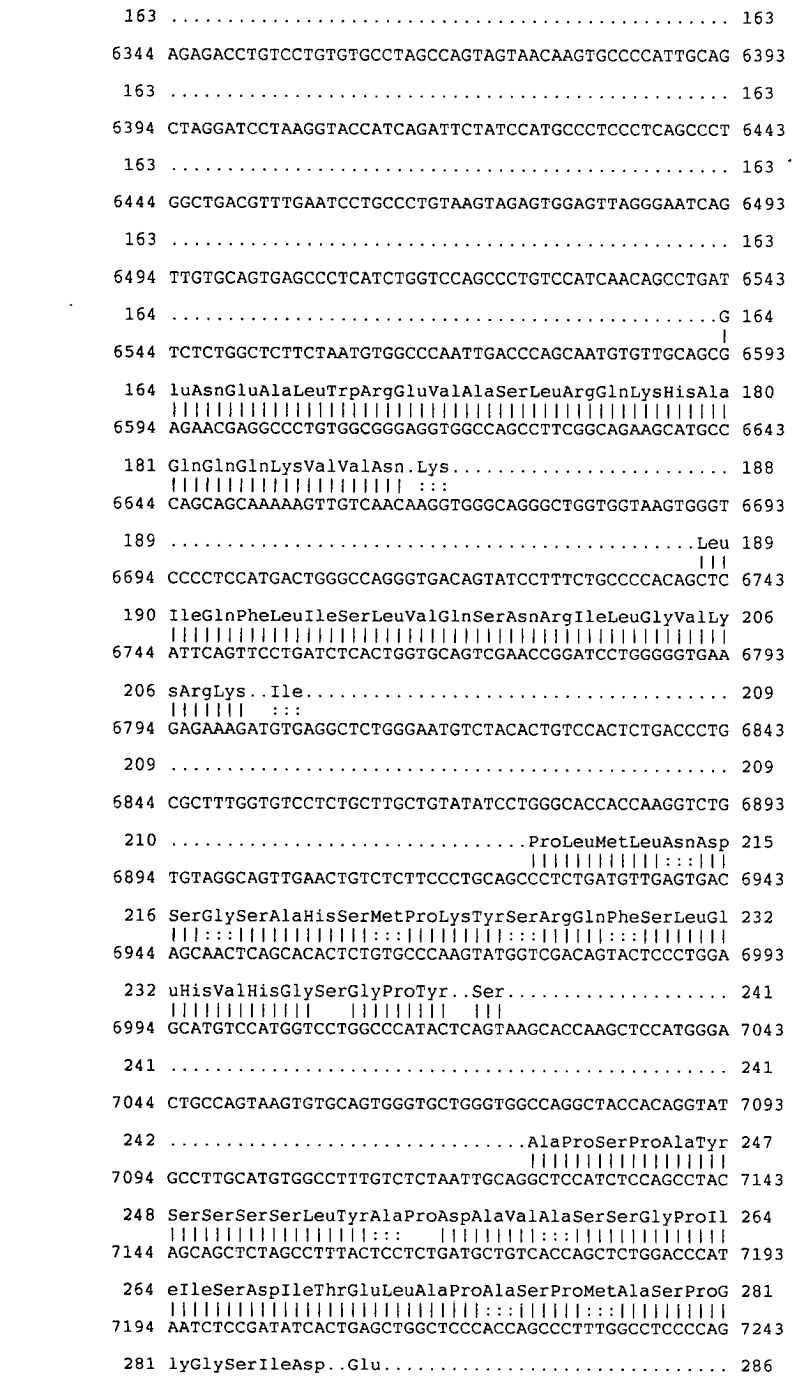
```
  381 ..................................................... 381
51208 GGTGCAGCCTGGGGGGTACAGTCAAGGGGAGCCCTTCTCCCACAGGAGGG 51159
  381 ..................................................... 381
51158 CATTGGGGTGTGGGGCCTGGGGCACTGGTTCAGGTACCGCCTTATCCCGG 51109
  382 ......AsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAs 396
          ||||||||||||||||||||||||||||||||||||||||||||||||
51108 GCCAGGAATGAGCTCAGTGACCACTTGGATGCTATGGACTCCAACCTGGA 51059
  396 pAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThrSerA 413
      ||||||||||||||||||||||||||||||||||||||||||||||||||
51058 TAACCTGCAGACCATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTG 51009
  413 la................................................... 413
      ||
51008 CCCTGCTGGACGTGAGTGGAGCCCCGCGGCCCCGCCTCCCCGCCCCGCCT 50959
  413 ..................................................... 413
50958 CCCCGCGCATCCCCGCCTCCCCGCCTCCCCGCGCCGCATCCCCGCTTCCC 50909
  413 ..................................................... 413
50908 CGCGCCGCATCCCCGCATCCCCGCGCCGCCTCCCCGCCTCCCCGCGCCGC 50859
  413 ..................................................... 413
50858 CTCCCCGCCTCCCCGCTTCCCCGCCCCGCCCCCGGGTGCTGTTCTGACTT 50809
  414 ......LeuLeu..AspLeuPheSerProSerValThrValProAspMet 427
          ||||||  :::||||||||||||||||||||||||||||||||||||
50808 CCCTCCCTCCTCCGCAGCTGTTCAGCCCCTCGGTGACCGTGCCCGACATG 50759
  428 SerLeuProAspLeuAspSerSerLeuAla..................... 437
      ||||||||||||||||||||||||||||||
50758 AGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTGTGCGTAGGCGGGCGGG 50709
  437 ..................................................... 437
50708 GGGTGAGGGGGAACGAGACCAGCGGGAGTGCTCACAATACCGTCTCCACC 50659
  438 ..Ser.IleGlnGluLeuLeuSerProGlnGluProProArgProProGl 453
        ::: ||||||||||||||||||||||||||||||||||||||||||||||
50658 CCACAGATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCCGA 50609
  453 uAlaGluAsnSerSerProAsp.Ser........................ 461
      |||||||||||||||||||||| :::
50608 GGCAGAGAACAGCAGCCCGGATTCAGGTGAGCCAAGTCCCACCGGCCCCA 50559
  462 .....................................GlyLy 463
                                          |||||
50558 CCTTTGCCCCCAACCCCCCACCGCCTTGACACCCCCACCCCCGCAGGGAA 50509
  463 sGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlyS 480
      ||||||||||||||||||||||||||||||||||||||||||||||||||
50508 GCAGCTGGTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCT 50459
  480 erValAspThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu 496
      ||||||||||||||||||||||||||||||||||||||||||||||||||
50458 CCGTGGACACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAG 50409
  497 GlySerTyrPheSerGluGlyAspGlyPheAlaGluAspProThrIleSe 513
      ||||||||||||||||||||||||||||||||||||||||||||||||||
50408 GGCTCCTACTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTC 50359
  513 rLeuLeuThrGlySerGluProProLysAlaLysAspProThrValSer 529
      ||||||||||||||||||||||||||||||||||||||||||||||||||
50358 CCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 50310
```

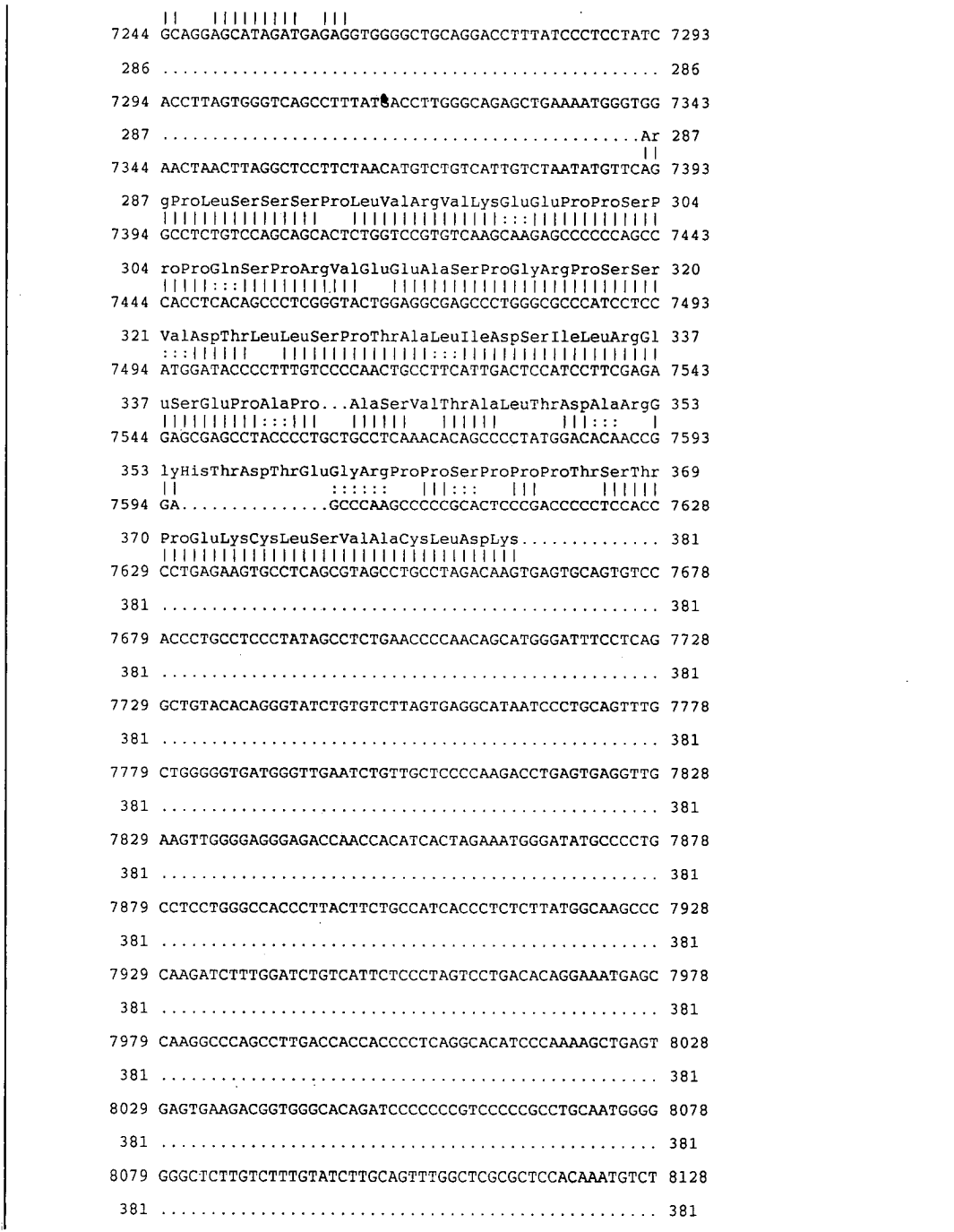
```
seq_name: gb_ov:XELHSF

seq_documentation_block:
LOCUS       XELHSF       1657 bp    mRNA            VRT       16-MAY-1996
DEFINITION  Xenopus laevis heat shock factor 1 (XHSF1) mRNA, complete cds.
ACCESSION   L36924
VERSION     L36924.1  GI:558067
KEYWORDS    heat shock factor.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
            Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
            Xenopus.
REFERENCE   1  (bases 1 to 1657)
  AUTHORS   Stump,D.G., Landsberger,N. and Wolffe,A.P.
  TITLE     The cDNA encoding Xenopus laevis heat-shock factor 1 (XHSF1):
            nucleotide and deduced amino-acid sequences, and properties of the
            encoded protein
  JOURNAL   Gene 160 (2), 207-211 (1995)
  MEDLINE   95369690
REFERENCE   2  (bases 1 to 1657)
  AUTHORS   Landsberger,N. and Wolffe,A.P.
  TITLE     Role of chromatin and Xenopus laevis heat shock transcription
            factor in regulation of transcription from the X. laevis hsp70
            promoter in vivo
  JOURNAL   Mol. Cell. Biol. 15 (11), 6013-6024 (1995)
  MEDLINE   96025982
FEATURES             Location/Qualifiers
     source          1. .1657
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /cell_line="A6"
     5'UTR           1. .226
     mRNA            1. .1657
     gene            227. .1582
                     /gene="XHSF1"
     CDS             227. .1582
                     /gene="XHSF1"
                     /codon_start=1
                     /product="heat shock factor"
                     /protein_id="AAA99999.1"
                     /db_xref="GI:558068"
                     /translation="MDPHGTCGGSNVPAFLAKLWTLVEDPDTDPLICWSPEGNSFHVF
                     DQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVHIEQGGLVKPERDDTEFQHPYFI
                     RGQEQLLENIKRKVNTMSATKSDEVKVRQDSVGKLISDVQSMKGKQESIDGRLLSMKH
                     ENEALWREVASLRQKHTQQQKVVNKLIQFLVSLVQSNRILGVKRKIPLMLNDSSTGHS
                     PPKYSRQYSLEHVPSSTSYPVSGFTDSSAGPIISDVTELPESSPSPSPCPSLEASPSP
                     VILIKTEPLTPSQSPEQSPAPPKLDDTPISPSTFIDSILLETETSVCPGGNKNDEMSE
                     SHPPEPCLSVACLDNISLSRQMSEVSRLFPTSCSSVPGRAEPPGLDMAVAELNDHVDN
                     IDFNLDTLQNLLNGQSFSVDTSALMDLFSPSLGIPDLSLPDPDSSLASVSSTPIYTCV
                     V"
     3'UTR           1583. .1657
BASE COUNT      400 a    467 c    409 g    381 t
ORIGIN


alignment_scores:
            Quality: 1406.50             Length:    468
              Ratio:  3.896               Gaps:      8
 Percent Similarity:  77.137   Percent Identity: 61.752

alignment_block:
 US-09-304-121-2 x XELHSF   ..

 Align seg 1/1  to: XELHSF  from: 1  to: 1657

       8 GlyAlaAlaGlyProSerAsnValProAlaPheLeuThrLysLeuTrpTh 24
         |||:::::::|||   |||||||||||||||||||||||:::|||||||||||
     239 GGGACTTGTGGGGGCAGCAATGTACCTGCTTTCCTTGCCAAGCTCTGGAC 288

      24 rLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerProSerG 41
         |||||||:::|||||||||||||   |||||||||||||||||||:::|
   .  289 TTTGGTGGAAGACCCAGACACAGACCCTTTAATCTGCTGGAGCCCAGAAG 338

      41 lyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGluValLeu 57
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     339 GGAACAGTTTCCATGTGTTTGACCAGGGACAGTTTGCCAAAGAGGTTCTT 388

      58 ProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGlnLeuAs 74
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     389 CCAAAGTATTTCAAGCACAACAACATGGCCAGCTTTGTACGGCAACTCAA 438

      74 nMetTyrGlyPheArgLysValValHisIleGluGlnGlyGlyLeuValL 91
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     439 TATGTATGGGTTTCGGAAGGTTGTCCATATTGAACAAGGGGGCCTGGTGA 488

      91 ysProGluArgAspAspThrGluPheGlnHisProCysPheLeuArgGly 107
         ||||||||||||||||||||||||||||||||||   |||:::||||||
     489 AACCTGAAAGAGATGACACAGAGTTCCAGCACCCCTACTTCATCCGTGGT 538

     108 GlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerValSerTh 124
         ||||||||||||||||||||||||||||||||||:::::::::|||::
     539 CAGGAGCAACTTCTGGAGAACATCAAGCGCAAAGTCAACACCATGTCTGC 588

     124 rLeuLysSerGluAspIleLysIleArgGlnAspSerValThrLysLeuL 141
         :   ||||||:::::::::|||:::|||||||||||||||   ||||||:
     589 AACAAAGAGCGACGAGGTGAAGGTGCGACAGGACAGTGTTGGCAAGTTGA 638

     141 euThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAspSerLys 157
         :::::|||||||||   ||||||||||||||||   :::|||:::::::
     639 TATCAGATGTACAGAGCATGAAGGGCAAGCAGGAGAGCATTGATGGCCGG 688

     158 LeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluValAlaSe 174
         |||||||:::||||||||||||||||||||||||||||||||||||||||
     689 CTACTCTCCATGAAACATGAGAACGAGGCCCTGTGGAGGGAGGTTGCGAG 738

     174 rLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLysLeuIleG 191
         ||||||||||||||||:::||||||||||||||||||||||||||||||||
     739 CCTGAGGCAGAAGCACACTCAGCAGCAGAAAGTGGTCAATAAGCTCATCC 788

     191 lnPheLeuIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArg 207
         ||||||||:::||||||||||||||||||||||||||||||||||||||||
     789 AGTTCCTGGTCTCATTGGTTCAGTCAAACCGGATCCTGGGGGTAAAGAGG 838

     208 LysIleProLeuMetLeuAsnAspSerGlySerAlaHisSerMetProLy 224
         ||||||||||||||||||||||||||:::::::::|||||   |||||
     839 AAAATCCCCCTCATGCTGAATGACAGTAGTACAGGCCACTCCCCTCCCAA 888

     224 sTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProTyrS 241
         ||||||||||||||:::|||||||||||||||   :::|||      |||
     889 GTACAGTCGTCAGTACTCATTGGAGCACGTGCCCAGCAGCACCTCCTAC. 937

     241 erAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAspAla 257
              |||      ::::::::::::|||
     938 .....CCAGTCTCGGGATTCACAGACTCC..................... 961

     258 ValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaProAl 274
               |||:::|||||||||||||||||||:::|||||||||      ::
     962 ......TCGGCAGGACCCATTATATCTGATGTGACGGAACTCCCAGAGTC 1005

     274 aSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuSerS 291
         :||||||      ||||||      |||::::::      |||
    1006 CAGCCCCTCCCCCTCTCCGTGTCCCAGTTTAGAGGCCAGCCCC....... 1048

     291 erSerProLeuValArgValLysGluGluProProSerProProGlnSer 307
           ||||||:::::   :::|||   ||||||   :::|||   ||||||
    1049 ..AGTCCAGTGATTCTGATCAAGACGGAGCCCCTCACTCCATCCCAGAGC 1096

     308 ProArgValGluGluAlaSerProGlyArgProSerSerValAspThrLe 324
         |||      |||   ||||||:::   |||:::      ||||||
    1097 CCT.........GAGCAGAGTCCTGCGCCCCCTAAACTGGACGACACCCC 1137

     324 uLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluProA 341
          :::|||||:::::::::|||||||||||||||   |||:::|||
    1138 AATATCGCCCTCCACTTTCATAGACTCAATTCTGCTGGAGACTGAG.... 1183
```

```
     341 laProAlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAsp... 356
                    |||::::::      :::   :::   :::|||
    1184 .............ACCAGTGTGTGTCCCGGGGGCAATAAAAATGATGAA 1219

     357 ...ThrGluGlyArgProProSerProProProThrSerThrProGluLy 372
           :::|||:::::::||||||:::|||
    1220 ATGTCAGAGTCGCACCCCCCTGAGCCC..................... 1246

     372 sCysLeuSerValAlaCysLeuAspLys..................... 381
          |||||||||||||||||||||||:::
    1247 .TGCCTGAGTGTCGCCTGTCTAGACAACATCTCTTTGAGCCGTCAGATGT 1295

     381 ................................................. 381

    1296 CTGAGGTCTCTCGTCTGTTCCCCACGTCCTGTTCCTCAGTGCCGGGCAGA 1345

     382 ............................AsnGluLeuSerAspHisLe 388
                                      ||||||:::||||||::
    1346 GCGGAGCCCCCCGGGTTAGATATGGCTGTGGCGGAGCTAAATGACCATGT 1395

     388 uAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerH 405
         :|||   :::|||   |||||||||:::||||||:::::::|||:::::::
    1396 GGACAACATTGACTTCAACCTGGACACCCTGCAGAACCTTCTCAATGGCC 1445

     405 isGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSer 421
         :::::|||||||||||||||||||||||||:::|||||||||||||||||
    1446 AGAGCTTCAGTGTAGATACCAGTGCTCTCATGGATCTCTTCAGTCCGTCC 1495

     422 ValThrValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSe 438
         :::   :::|||||||:::|||||||||||   ||||||||||||||||
    1496 TTGGGAATCCCAGATCTGTCCCTCCCAGACCCCGACTCCAGCCTCGCCAG 1545

     438 rIle 439
         |:::
    1546 TGTG 1549
```

```
seq_name: gb_ro:AF059275

seq_documentation_block:
LOCUS       AF059275    11395 bp    DNA             ROD       01-DEC-1998
DEFINITION  Mus musculus heat shock transcription factor 1 (Hsf1) gene, partial
            cds.
ACCESSION   AF059275
VERSION     AF059275.1  GI:3126917
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 11395)
  AUTHORS   Zhang,Y., Koushik,S., Dai,R. and Mivechi,N.F.
  TITLE     Structural organization and promoter analysis of murine heat shock
            transcription factor-1 gene
  JOURNAL   J. Biol. Chem. 273 (49), 32514-32521 (1998)
  MEDLINE   99047622
REFERENCE   2  (bases 1 to 11395)
  AUTHORS   Zhang,Y., Koushik,S.V., Dai,R. and Mivechi,N.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1998) Institute of Molecular Medicine & Genetics,
            1120 15th Street, Augusta, GA 30912-2630, USA
FEATURES             Location/Qualifiers
     source          1. .11395
                     /organism="Mus musculus"
                     /strain="129SVJ"
                     /db_xref="taxon:10090"
                     /tissue_type="liver"
                     /sex="female"
                     /dev_stage="4-8 weeks"
                     /clone="3-1"
     intron          <1. .4943
                     /gene="Hsf1"
                     /number=1
     gene            <1. .9886
                     /gene="Hsf1"
     mRNA            join(<4944. .5052,5427. .5563,5810. .5934,6592. .6667,
                     6741. .6802,6925. .7021,7126. .7262,7394. .7663,8918. .9023,
                     9101. .9166,9237. .9306,9387. .9886)
                     /gene="Hsf1"
                     /product="heat shock transcription factor 1"
     CDS             join(<4944. .5052,5427. .5563,5810. .5934,6592. .6667,
                     6741. .6802,6925. .7021,7126. .7262,7394. .7663,8918. .9023,
                     9101. .9166,9237. .9306,9387. .9592)
                     /gene="Hsf1"
                     /codon_start=1
                     /product="heat shock transcription factor 1"
                     /protein_id="AAC80425.1"
                     /db_xref="GI:3126918"
                     /translation="SGNSFHVFDQGQFAKEVLPKYFKHNNMASFVRQLNMYGFRKVVH
                     IEQGGLVKPERDDTEFQHPCFLRGQEQLLENIKRKVTSVSTLKSEDIKIRQDSVTRLL
                     TDVQLMKGKQECMDSKLLAMKHENEALWREVASLRQKHAQQQKVVNKLIQFLISLVQS
                     NRILGVKRKIPLMLSDSNSAHSVPKYGRQYSLEHVHGPGPYSAPSPAYSSSSLYSSDA
                     VTSSGPIISDITELAPTSPLASPGRSIDERPLSSSTLVRVKQEPPSPPHSPRVLEASP
                     GRPSSMDTPLSPTAFIDSILRESEPTPAASNTAPMDTTGAQAPALPTPSTPEKCLSVA
                     CLDKNELSDHLDAMDSNLDNLQTMLTSHGFSVDTSALLDLFSPSVTMPDMSLPDLDSS
                     LASIQELLSPQEPPRPIEAENSNPDSGKQLVHYTAQPLFLLDPDAVDTGSSELPVLFE
                     LGESSYFSEGDDYTDDPTISLLTGTEPHKAKDPTVS"
     exon            4944. .5052
                     /gene="Hsf1"
                     /number=2
     intron          5053. .5426
                     /gene="Hsf1"
                     /number=2
     exon            5427. .5563
                     /gene="Hsf1"
                     /number=3
     intron          5564. .5809
                     /gene="Hsf1"
                     /number=3
     exon            5810. .5934
                     /gene="Hsf1"
                     /number=4
     intron          5935. .6591
                     /gene="Hsf1"
                     /number=4
     exon            6592. .6667
                     /gene="Hsf1"
                     /number=5
     intron          6668. .6740
                     /gene="Hsf1"
                     /number=5
     exon            6741. .6802
                     /gene="Hsf1"
                     /number=6
     intron          6803. .6924
                     /gene="Hsf1"
                     /number=6
     exon            6925. .7021
                     /gene="Hsf1"
                     /number=7
     intron          7022. .7125
                     /gene="Hsf1"
                     /number=7
     exon            7126. .7262
                     /gene="Hsf1"
                     /number=8
     intron          7263. .7393
                     /gene="Hsf1"
                     /number=8
     exon            7394. .7663
                     /gene="Hsf1"
                     /number=9
     intron          7664. .8917
                     /gene="Hsf1"
                     /number=9
     exon            8918. .9023
                     /gene="Hsf1"
```

```
                     /number=10
     intron          9024. .9100
                     /gene="Hsf1"
                     /number=10
     exon            9101. .9166
                     /gene="Hsf1"
                     /note="isoform"
                     /number=11
     intron          9167. .9236
                     /gene="Hsf1"
                     /number=11
     exon            9237. .9306
                     /gene="Hsf1"
                     /number=12
     intron          9307. .9386
                     /gene="Hsf1"
                     /number=12
     exon            9387. .9886
                     /gene="Hsf1"
                     /number=13
BASE COUNT     2685 a   3044 c   2670 g   2996 t
ORIGIN


alignment_scores:
           Quality: 1393.00             Length:   1557
             Ratio:  2.996                Gaps:     21
 Percent Similarity:  29.865   Percent Identity: 27.617

alignment_block:
 US-09-304-121-2 x AF059275   ..

 Align seg 1/1  to: AF059275  from: 1  to: 11395

      40 SerGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGluVa 56
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4944 AGTGGGAACAGCTTCCACGTGTTTGACCAGGGCCAGTTTGCCAAGGAGGT 4993

      56 lLeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGlnL 73
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4994 GCTGCCCAAGTACTTCAAGCACAACAACATGGCTAGCTTCGTGCGGCAGC 5043

      73 euAsnMet..................................... 75
         ||||||||
    5044 TCAACATGTGTAAGTACCTCTTTCAGGATGCTATGCTTGCTTCACTGGGT 5093

      75 ..................................................... 75

    5094 GTGGAGAAATAGACCCCAGCCTATTTCTAAGAGATGGAGGCCTGGGTTGG 5143

      75 ..................................................... 75

    5144 CTGAAGAAGGGTCCTCATGAGTTTGAGCCTGAATTCTCCAGGGAAGAGTG 5193

      75 ..................................................... 75

    5194 TAGGATATGAGGCCCTAACCTAATGGCCACACTTCCATTTACAAAAAGGG 5243

      75 ..................................................... 75

    5244 AAGCTGTTCGGGGACAAAGTCTTTTCCAACCAAGAATAACTTGCTATGTA 5293

      75 ..................................................... 75

    5294 TGGTAAATGGACACCAGTCAGACTCTCGTACAGTTTTCTAGGAGGCCCCT 5343

      75 ..................................................... 75

    5344 TGCCTGCCCTAGGCCATGGCTAGGCTGTTGAAGGTCTGGTCCCTGGCCCT 5393

      76 ..............................Tyr..GlyPheArgLysVal 81
                                          ||||||||||||||||
    5394 CAAACAGAACTTCCCCCACCCCCATATACACAGATGGCTTCCGAAAAGTA 5443

      82 ValHisIleGluGlnGlyGlyLeuValLysProGluArgAspAspThrGl 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5444 GTCCACATTGAGCAGGGTGGCCTGGTCAAGCCTGAGAGAGATGACACCGA 5493

      98 uPheGlnHisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnI 115
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5494 GTTCCAGCATCCTTGTTTCTTGCGTGGACAGGAACAGCTCCTTGAGAACA 5543

     115 leLysArgLysValThrSer..................................... 121
         ||||||||||||||||||||
    5544 TCAAGAGGAAAGTGACCAGCGTGAGTCACCAGCCCTGAGTACAGCCCCTC 5593

     121 ..................................................... 121

    5594 CCTGCCCTAAATATTCCTGCAGGCTATGGGCTGGGAGCAGTATCTAGGTG 5643

     121 ..................................................... 121

    5644 GTGCTGCAGGACTAGCCTAAAAGCATAGCCTGTGCTCTCTGTCTCTGACC 5693

     121 ..................................................... 121

    5694 CCCTCCCCCTTCCCAGTGTTTCCCAGGCCCTCCTCATTCCCTGGATGCCT 5743

     121 ..................................................... 121

    5744 AGGTGAAGCACAGGAGCCTACCACACTCCAGAGCCACTTTCGACTCTTGT 5793

     122 ................ValSerThrLeuLysSerGluAspIleLysIleA 133
                          ||||||||||||||||||||||||||||||||||
    5794 ATGCCCCACCCCTTAGGTGTCCACCCTGAAGAGTGAGGACATAAAAATAC 5843

     133 rgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGly 149
         ||||||||||||||||:::||||||||||||||||||||||||||||||
    5844 GCCAGGACAGTGTCACCCGGCTGTTGACAGATGTGCAGCTGATGAAGGGG 5893

     150 LysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHis........ 163
         ||||||||||||||||||||||||||||||||||||:::
    5894 AAACAGGAGTGTATGGACTCCAAGCTCCTGGCCATGAAGCAGTAGGTCCC 5943

     163 ..................................................... 163

    5944 AAGGCTCAGGAGGGTGGCCAGGTTGCTTTGGGTCTCCTACCCGATCCTGC 5993

     163 ..................................................... 163

    5994 AGAGCCATTCAAGCCTCCTTCCCCCTCCCAGCCTGCCTTCATTGTTTCTA 6043

     163 ..................................................... 163

    6044 TGGAAACCCTGTCTTGTACATCTCCAGACACAGAGTGGGGTTGGAGGTCT 6093

     163 ..................................................... 163

    6094 CCTACTGCACCCACTTCTCACTCCCCTGCTGGAATTCACATTTTCAGGAT 6143

     163 ..................................................... 163

    6144 TTGAGCACACAAAGCCCATTTCCAGGTGCCTCCCTAAAAACCAGGCCTGA 6193

     163 ..................................................... 163

    6194 CAGGAGCTGACCACAACAGTCCTTGGGCCAGACCTCTCTACTGTTCTCTA 6243

     163 ..................................................... 163

    6244 GTCACAATGACTCAACTCGCCCTTTATCACTCACTCAGCACAGTTCACAG 6293

     163 ..................................................... 163

    6294 CTGCCTTCTACCCCAAAACTGAGGGTAATCTGTAGTTTAGGATCTTTCCC 6343
```


163 163
6344 AGAGACCTGTCCTGTGTGCCTAGCCAGTAGTAACAAGTGCCCCATTGCAG 6393
163 163
6394 CTAGGATCCTAAGGTACCATCAGATTCTATCCATGCCCTCCCTCAGCCCT 6443
163 163
6444 GGCTGACGTTTGAATCCTGCCCTGTAAGTAGAGTGGAGTTAGGGAATCAG 6493
163 163
6494 TTGTGCAGTGAGCCCTCATCTGGTCCAGCCCTGTCCATCAACAGCCTGAT 6543
164 .....G 164
6544 TCTCTGGCTCTTCTAATGTGGCCCAATTGACCCAGCAATGTGTTGCAGCG 6593
164 luAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla 180
6594 AGAACGAGGCCCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCC 6643
181 GlnGlnGlnLysValValAsn.Lys..... 188
6644 CAGCAGCAAAAAGTTGTCAACAAGGTGGGCAGGGCTGGTGGTAAGTGGGT 6693
189 .....Leu 189
6694 CCCCTCCATGACTGGGCCAGGGTGACAGTATCCTTTCTGCCCCACAGCTC 6743
190 IleGlnPheLeuIleSerLeuValGlnSerAsnArgIleLeuGlyValLy 206
6744 ATTCAGTTCCTGATCTCACTGGTGCAGTCGAACCGGATCCTGGGGGTGAA 6793
206 sArgLys..Ile..... 209
6794 GAGAAAGATGTGAGGCTCTGGGAATGTCTACACTGTCCACTCTGACCCTG 6843
209 209
6844 CGCTTTGGTGTCCTCTGCTTGCTGTATATCCTGGGCACCACCAAGGTCTG 6893
210 .....ProLeuMetLeuAsnAsp 215
6894 TGTAGGCAGTTGAACTGTCTCTTCCCTGCAGCCCTCTGATGTTGAGTGAC 6943
216 SerGlySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGl 232
6944 AGCAACTCAGCACACTCTGTGCCCAAGTATGGTCGACAGTACTCCCTGGA 6993
232 uHisValHisGlySerGlyProTyr..Ser..... 241
6994 GCATGTCCATGGTCCTGGCCCATACTCAGTAAGCACCAAGCTCCATGGGA 7043
241 241
7044 CTGCCAGTAAGTGTGCAGTGGGTGCTGGGTGGCCAGGCTACCACAGGTAT 7093
242 .....AlaProSerProAlaTyr 247
7094 GCCTTGCATGTGGCCTTTGTCTCTAATTGCAGGCTCCATCTCCAGCCTAC 7143
248 SerSerSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIl 264
7144 AGCAGCTCTAGCCTTTACTCCTCTGATGCTGTCACCAGCTCTGGACCCAT 7193
264 eIleSerAspIleThrGluLeuAlaProAlaSerProMetAlaSerProG 281
7194 AATCTCCGATATCACTGAGCTGGCTCCCACCAGCCCTTTGGCCTCCCCAG 7243
281 lyGlySerIleAsp..Glu..... 286


7244 GCAGGAGCATAGATGAGAGGTGGGGCTGCAGGACCTTTATCCCTCCTATC 7293
286 286
7294 ACCTTAGTGGGTCAGCCTTTATGACCTTGGGCAGAGCTGAAAATGGGTGG 7343
287 .....Ar 287
7344 AACTAACTTAGGCTCCTTCTAACATGTCTGTCATTGTCTAATATGTTCAG 7393
287 gProLeuSerSerSerProLeuValArgValLysGluGluProProSerP 304
7394 GCCTCTGTCCAGCAGCACTCTGGTCCGTGTCAAGCAAGAGCCCCCCAGCC 7443
304 roProGlnSerProArgValGluGluAlaSerProGlyArgProSerSer 320
7444 CACCTCACAGCCCTCGGGTACTGGAGGCGAGCCCTGGGCGCCCATCCTCC 7493
321 ValAspThrLeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGl 337
7494 ATGGATACCCCTTTGTCCCCAACTGCCTTCATTGACTCCATCCTTCGAGA 7543
337 uSerGluProAlaPro...AlaSerValThrAlaLeuThrAspAlaArgG 353
7544 GAGCGAGCCTACCCCTGCTGCCTCAAACACAGCCCCTATGGACACAACCG 7593
353 lyHisThrAspThrGluGlyArgProProSerProProProThrSerThr 369
7594 GA.....GCCCAAGCCCCCGCACTCCCGACCCCCTCCACC 7628
370 ProGluLysCysLeuSerValAlaCysLeuAspLys..... 381
7629 CCTGAGAAGTGCCTCAGCGTAGCCTGCCTAGACAAGTGAGTGCAGTGTCC 7678
381 381
7679 ACCCTGCCTCCCTATAGCCTCTGAACCCCAACAGCATGGGATTTCCTCAG 7728
381 381
7729 GCTGTACACAGGGTATCTGTGTCTTAGTGAGGCATAATCCCTGCAGTTTG 7778
381 381
7779 CTGGGGGTGATGGGTTGAATCTGTTGCTCCCCAAGACCTGAGTGAGGTTG 7828
381 381
7829 AAGTTGGGGAGGGAGACCAACCACATCACTAGAAATGGGATATGCCCCTG 7878
381 381
7879 CCTCCTGGGCCACCCTTACTTCTGCCATCACCCTCTCTTATGGCAAGCCC 7928
381 381
7929 CAAGATCTTTGGATCTGTCATTCTCCCTAGTCCTGACACAGGAAATGAGC 7978
381 381
7979 CAAGGCCCAGCCTTGACCACCACCCCTCAGGCACATCCCAAAAGCTGAGT 8028
381 381
8029 GAGTGAAGACGGTGGGCACAGATCCCCCCCGTCCCCCGCCTGCAATGGGG 8078
381 381
8079 GGGCTCTTGTCTTTGTATCTTGCAGTTTGGCTCGCGCTCCACAAATGTCT 8128
381 381

```
8129 GGGGTCGCCCGCCTCTTCCCCTGCCCCTCTTCCTTTCTGCATGGCCGAAT 8178
 381 ..................................................... 381
8179 CCAACCAGGGTTATCGCTGCCCCACTTGGGAAGGAAGAGGACTCCGCCAG 8228
 381 ..................................................... 381
8229 GCCCCCTCCCACAGCCATGGACCAGTTCCCAACCTGGGCATAGTACTTGC 8278
 381 ..................................................... 381
8279 CCTGTCTGGCGAAGAGCACCAGGGGATGGTCTCACAGGCTCTGCATCTCC 8328
 381 ..................................................... 381
8329 ACTGACTACCCTTTGCACAAGCCAATGGCTGTAATAGGGACAAGTGTCCT 8378
 381 ..................................................... 381
8379 GCCTGCCCCAGCTACCAGTACAGGCTGCTACAATGAGCTTCTTGTCAGAG 8428
 381 ..................................................... 381
8429 TAGGGTAGGCCCTATGGCACCTACAGAGCCATTTTAACTGCCTTGCCCAG 8478
 381 ..................................................... 381
8479 GCTCAGGTGTGTCCATGGCCTACGGCCTTCTGGTCTCCCCATGCCAAGTA 8528
 381 ..................................................... 381
8529 CAGTCCTGCTTCCTCTCCTTACTCTCTTCCTGGTGCACTTGCAAGACTGT 8578
 381 ..................................................... 381
8579 CAAACCTGAAAAATCTGGCCGCCTATCCATTTGCAGACCAATGTCCCTTG 8628
 381 ..................................................... 381
8629 TCTTTTTTCTCCACCTCCCTCTTGGCCCAGCCTTAGTAACTACTAAAGCC 8678
 381 ..................................................... 381
8679 TCCTAAGATGTCCTGCCCTGGGCCATGCTGTATCCCACCTCAGCAGCCAT 8728
 381 ..................................................... 381
8729 GGAAGGCAAGCTTACTACTTGTCCTTTCGGCATCTTCCCAGAGATACACA 8778
 381 ..................................................... 381
8779 CAGTCCCCCCAGCCCATACTGGCTATGCAGAAGGACCTTGCTTGGGGGAC 8828
 381 ..................................................... 381
8829 TGCAACCAAAGGGGAGCCTTTCTCCAGCAGGAGTGTGGCTGGGTGGGAGC 8878
 382 .....................................AsnGluLeuS 385
                                          ||||||||||
8879 CTGGGGCACTAGTTCACCTGCCCACTTGCCTTGGGCCAGGAACGAGCTAA 8928
 385 erAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMet 401
      |||||||||||||||||||||||||||||||||||||||||||||||||
8929 GTGATCACCTGGATGCCATGGACTCCAACCTGGACAACCTGCAGACCATG 8978
 402 LeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAsp..... 416
      |||:::|||||||||||||||||||||||||||||||||||||||||
8979 CTGACAAGCCACGGCTTCAGTGTGGACACCAGTGCCCTGCTGGACGTGAG 9028
 416 ..................................................... 416
9029 TGTGGTCATCCCTACCCACCCTGCTCCATCCTGCCCACAAGCCAGCCCTG 9078
```

```
 417 ...................Leu..PheSerProSerValThrValProA 426
                                ||||||||||||||||||:::||||
9079 ACTTCCCTCCCTCCTCCTGCAGCTGTTCAGCCCCTCGGTGACCATGCCCG 9128
 426 spMetSerLeuProAspLeuAspSerSerLeuAla..............  437
      |||||||||||||||||||||||||||||||||
9129 ACATGAGCCTGCCTGACCTGGACAGCAGCCTGGCCAGCGTGCGTAGGCGG 9178
 437 ..................................................... 437
9179 GCAGGGTGGGGGGCAGAGGGGGCCATCAACAACCCTCACTGTGTTCCCTG 9228
 438 ....Ser.IleGlnGluLeuLeuSerProGlnGluProProArgProPro 452
          ::: ||||||||||||||||||||||||||||||||||||||||||
9229 TCCCACAGATTCAGGAGCTTCTGTCTCCACAAGAGCCTCCCAGGCCTATT 9278
 453 GluAlaGluAsnSerSerProAspSerGly..................... 462
      ||||||||||||||:::|||||||||||
9279 GAGGCAGAGAACAGTAACCCCGACTCAGGTGAGAGCTAAGTTTTTTTGCT 9328
 462 ..................................................... 462
9329 ACCCCCAGCCCTCACTTCCAGATTGCCCACCCCACCATCCAAACCCCATT 9378
 463 ........Lys..GlnLeuValHisTyrThrAlaGlnProLeuPheLeuL 476
             :::  ||||||||||||||||||||||||||||||||||||||||
9379 CCCCACAGGAAAGCAGCTGGTGCACTACACGGCTCAGCCTCTGTTCCTGC 9428
 476 euAspProGlySerValAspThrGlySerAsnAspLeuProValLeuPhe 492
      ||||||||   :::||||||||||||||||:::::|||||||||||||||
9429 TGGATCCTGATGCTGTGGACACAGGGAGCAGTGAGCTGCCTGTGCTCTTT 9478
 493 GluLeuGlyGluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAs 509
      ||||||||||||:::|||||||||||||||||||||||   ::::::::::||
9479 GAGCTGGGGGAGAGCTCCTACTTCTCTGAGGGGGATGACTACACGGATGA 9528
 509 pProThrIleSerLeuLeuThrGlySerGluProProLysAlaLysAspP 526
      ||||||||||||||||||||||||||:::||||||   ||||||||||||
9529 TCCCACCATCTCTCTTCTGACAGGCACTGAACCCCATAAAGCCAAGGACC 9578
 526 roThrValSer 529
      ||||||||||
9579 CCACTGTCTCC 9589
```

```
seq_name: gb_ov:CHKHSF3B

seq_documentation_block:
LOCUS       CHKHSF3B     2366 bp    mRNA            VRT       17-JUN-1993
DEFINITION  Chicken mRNA sequence.
ACCESSION   L06125
VERSION     L06125.1  GI:289815
KEYWORDS    .
SOURCE      Gallus gallus (sub_species domesticus) cDNA to mRNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2366)
  AUTHORS   Nakai,A. and Morimoto,R.I.
  TITLE     Characterization of a novel chicken heat sock transcription factor,
            heat shock factor 3, suggests a new regulatory pathway
  JOURNAL   Mol. Cell. Biol. 13, 1983-1997 (1993)
  MEDLINE   93204945
FEATURES             Location/Qualifiers
     source          1. .2366
                     /organism="Gallus gallus"
                     /sub_species="domesticus"
                     /db_xref="taxon:9031"
BASE COUNT      719 a    498 c    531 g    618 t
ORIGIN


alignment_scores:
```

```
          Quality:  814.00            Length:    570
            Ratio:  2.353               Gaps:     18
Percent Similarity:  60.702   Percent Identity: 37.544

alignment_block:
 US-09-304-121-2 x CHKHSF3B   ..

 Align seg 1/1  to: CHKHSF3B  from: 1  to: 2366

      7 ProGlyAlaAlaGlyProSerAsnValProAlaPheLeuThrLysLeuTr 23
        |||::::::|||||          |||||||||||||||:::|||||||
     59 CCCGCCGGCGCCGGG.........GTGCCGGCCTTCCTCAGCAAGCTGTG 99

     23 pThrLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerProS 40
        |:::||||||::::::      ::::::   ||||||   ||||||   :
    100 GGCGCTGGTGGGCGAGGCGCCCAGCAACCAGCTTATCACCTGGAGCCAGA 149

     40 erGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGluVal 56
        ::|||:::||||||   |||:::|||:::   :::|||||||||||:::
    150 ATGGCCAGAGTTTCTTGGTGTTGGATGAACAGAGATTTGCAAAAGAGATT 199

     57 LeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGlnLe 73
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    200 CTGCCTAAGTACTTCAAGCACAACAACATGGCAAGCTTTGTCAGACAGCT 249

     73 uAsnMetTyrGlyPheArgLysValValHisIleGluGlnGlyGlyLeuV 90
        ||||||||||||||||||||||||||||||:::::   :::|||:::|
    250 AAACATGTATGGCTTCCGTAAAGTTGTCCATGTTGAC...TCTGGGATTG 296

     90 alLysProGluArgAspAsp...ThrGluPheGlnHisProCysPheLeu 105
        |||||   |||||||||      :::|||||||||||||||   |||
    297 TCAAACTGGAACGAGATGGTCTAGTTGAGTTTCAGCACCCGTATTTTAAG 346

    106 ArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerVa 122
        :::|||:::|||:::||||||||||:::||||||||||||         ||
    347 CAGGGTCGAGAGGACTTGTTGGAACACATTAAAAGGAAG.........GT 387

    122 lSerThrLeuLysSerGluAspIleLysIleArgGlnAspSerValThrL 139
        ||||:::   :::   |||:::   ||||||   |||:::::::::::::|
    388 TTCTTCTTCAAGACCTGAAGAAAACAAGATAAGCCAGGAGGATCTCTCCA 437

    139 ysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAsp 155
        ||:::::::::::::::|||   ::::::   |||||||||   ::::::
    438 AAATAATAAGTAGTGCTCAGAAAGTGGAGATTAAACAAGAGACTATTGAA 487

    156 SerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluVa 172
        |||:::|||   |||:::|||:::|||||||||:::|||||||||||||||
    488 TCTCGGTTGTCTGCTCTGAAAAGGGAGAATGAATCTCTTTGGAGGGAAGT 537

    172 lAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLysL 189
        ||||:::||||||   ||||||   :::|||||:::|||:::::::::|:
    538 GGCAGAACTGAGAGCAAAACACCTGAAACAACAGCAAGTTATTCGGAAGA 587

    189 euIleGlnPheLeuIleSerLeuValGlnSerAsnArgIleLeuGlyVal 205
        :::::||||||:::::::::|||||||||:::|||:::::::::::::::
    588 TTGTACAGTTTATTGTTACCTTGGTGCAGAATAACCAACTAGTGAGCCTA 637

    206 LysArgLysIleProLeuMetLeuAsnAspSerGlySerAlaHisSerMe 222
        |||||||||   ||||||:::||||||   :::|||
    638 AAACGCAAGAGGCCTCTACTTCTGAACACTAATGGA.............. 673

    222 tProLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyP 239
         |||      |||:::   |||:::   :::   |||         :::
    674 .CCTACAAAGTCGAATGTATTTCAG...CAAATTGTGAAAGAACCAGCAG 719

    239 roTyrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaPro 255
            :::         |||         :::::::::|||
    720 ACAATAATAACCATGTACCTCTCAACAGAACTGAGGGCTTAAAACAAAGA 769

    256 AspAlaValAlaSerSerGlyProIleIleSerAspIleThr........ 269
        :::   :::   |||:::      ||||||   |||:::|||
    770 GAACAAATT...TCAGATGACATCATTATTTATGATGTTACTGAGGATGT 816

    270 ............................GluLeuAlaProAlaSerP 276
                                    |||   ||||||   :::|
    817 GGCTGATGAGGAAAATACCATGGTTGATGAAGAAAATGCTCCCATTACAC 866

    276 roMetAlaSerProGlyGlySerIleAspGluArgProLeuSerSerSer 292
        ||   ::::::         :::   |||:::         |||   |||
    867 CAGAGACAAATGAAGATACCACCTCAGATTCTTCTAACTGCAGTCGTTCT 916

    293 ProLeuValArgVal............LysGluGluProProSerPro.. 304
        |||   :::   :::            :::|||||||      :::|||
    917 CCTGATATCGTAATTGTGGAAGATGATAATGAGGAAGAATATGCCCCTGT 966

    305 ..............................................P 305
                                                     |
    967 AATTCAAGGGGATAAAAGCACAGAATCAGTTGCTGTCTCAGCTAATGATC 1016

    305 roGlnSerProArgValGluGluAlaSerProGlyArgProSerSerVal 321
        ||   ||||||      :::::::::||||||         |||:::|||
   1017 CCCTCAGCCCCGTCAGTGACAGCACAAGTCCGCTCATGTCAAGCGCTGTA 1066

    322 Asp.....................ThrLeuLeuSerProThrAlaLeuIl 331
        :::                     |||       :::|||:::::::::::
   1067 CAGCTGAATAACCAGTCAACTTTAACTGCAGAAGACCCAGTCTCAATGAT 1116

    331 eAspSerIleLeuArgGluSerGluProAlaProAlaSerValThrAlaL 348
        :|||||||||||||:::|||:::                 :::::::::
   1117 GGATTCCATACTCAATGAAAATGGAGTAATTTCACAGAATATAAAT.... 1162

    348 euThrAspAlaArgGlyHisThrAspThrGluGlyArgProProSerPro 364

   1162 ..............................................  1162

    365 ProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLy 381
                                                   |||   ||
   1163 ....................................CTTCTTGGAAA 1173

    381 sAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnL 398
        |   ||||||   |||:::||||||:::::|||   :::|||::::::::
   1174 AGTTGAACTTCTGGATTATCTTGACAGTATCGACTGCAGTTTAGAGGACT 1223

    398 euGlnThrMetLeuSerSerHisGlyPheSerValAspThrSerAlaLeu 414
        ::|||:::||||||||:::::   ||||||:::|||   :::   |||
   1224 TCCAGGCCATGTTATCAGGACGACAGTTCAGCATAGATCCAGATCTGCTG 1273

    415 LeuAspLeuPheSerProSerValThrVal...ProAspMetSerLeuPr 430
        :::||||||||:::   ||||||   :::   |||         :::||
   1274 TTTGATCTTTTTACAAGCTCCGTGCAGATGAATCCCACAGATCATATCCC 1323

    430 oAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLeuSerProGlnG 447
        |:::         :::      :::|||:::         |||
   1324 TAATACCAAAATGGAGACAAAGGGAATAGAAACTACCAAGAGTAATGCTG 1373

    447 luProProArgProProGluAlaGluAsnSerSerProAspSerGlyLys 463
          |||          |||::::::   |||:::|||   |||   |||
   1374 GTCCAGCTGCTTCACAGGAAACACAAGTTTCTAAGCCTAAATCAGATAAA 1423

    464 GlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeu.........As 477
        ||||||:::::|||||||||   ||||||:::   :::         ::
   1424 CAACTCATACAGTATACTGCATTTCCGCTGCTCGCATTCCTTGACGGGAA 1473

    477 pProGlySer...ValAspThrGlySerAsn................... 486
        :|||||||||   |||:::::||||||:::
   1474 CCCAGGCTCCACTGTTGAGAGTGGGAGCTCCGCAACTGAAACTCCTTCCT 1523

    487 .....AspLeuProVal...LeuPheGluLeuGlyGluGlySerTyrPhe 500
             |||   |||:::   :::   ||||||   |||:::|||
   1524 CTGTGGACAAACCCCTGGAGGTGGATGAGCTCCTGGAGAGCAGCCTGGAT 1573

    501 SerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGl 517
           |||                  ||||||   |||   |||:::
   1574 CCCGAA.....................CCCACTCAGAGCAAGCTGGTGCG 1602

    517 ySerGluPro 520
```

```
           ||||||
   1603 CTTGGAGCCG 1612

seq_name: gb_pr1:HUMHSF2

seq_documentation_block:
LOCUS       HUMHSF2      2411 bp    mRNA            PRI       27-APR-1993
DEFINITION  Human heat shock factor 2 (HSF2) mRNA, complete cds.
ACCESSION   M65217
VERSION     M65217.1  GI:184404
KEYWORDS    heat shock factor 2.
SOURCE      Human, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2411)
  AUTHORS   Schuetz,T.J., Sheldon,L., Gallo,G.J., Tempst,P. and Kingston,R.E.
  TITLE     Isolation of a cDNA for HSF2: Evidence for two heat shock factor
            genes in humans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 6911-6915 (1991)
  MEDLINE   91334377
FEATURES             Location/Qualifiers
     source          1..2411
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HPB-ALL"
                     /cell_type="T-cells"
                     /tissue_type="lymphocyte"
     gene            89..1699
                     /gene="heat shock factor 2"
     CDS             89..1699
                     /gene="heat shock factor 2"
                     /codon_start=1
                     /product="HSF2"
                     /protein_id="AAA36017.1"
                     /db_xref="GI:184405"
                     /translation="MKQSSNVPAFLSKLWTLVEETHTNEFITWSQNGQSFLVLDEQRF
                     AKEILPKYFKHNNMASFVRQLNMYGFRKVVHIDSGIVKQERDGPVEFQHPYFKQGQDD
                     LLENIKRKVSSSKPEENKIRQEDLTKIISSAQKVQIKQETIESRLSELKSENESLWKE
                     VSELRAKHAQQQQVIRKIVQFIVTLVQNNQLVSLKRKRPLLLNTNGAQKKNLFQHIVK
                     EPTDNHHHKVPHSRTEGLKPRERISDDIIIYDVTDDNADEENIPVIPETNEDVISDPS
                     NCSQYPDIVIVEDDNEDEYAPVIQSGEQNEPARESLSSGSDGSSPLMSSAVQLNGSSS
                     LTSEDPVTMMDSILNDNINLLGKVELLDYLDSIDCSLEDFQAMLSGRQFSIDPDLLVD
                     LFTSSVQMNPTDYINNTKSENKGLETTKNNVVQPVSEEGRKSKSKPDKQLIQYTAFPL
                     LAFLDGNPASSVEQASTTASSEVLSSVDKPIEVDELLDSSLDPEPTQSKLVRLEPLTE
                     AEASEATLFYLCELAPAPLDSDMPLLDS"
BASE COUNT      734 a    458 c    497 g    722 t
ORIGIN


alignment_scores:
            Quality:  791.50              Length:    544
              Ratio:  2.335                 Gaps:     16
 Percent Similarity:  62.316    Percent Identity: 37.868

alignment_block:
 US-09-304-121-2 x HUMHSF2    ..

 Align seg 1/1  to: HUMHSF2  from: 1  to: 2411

      13 SerAsnValProAlaPheLeuThrLysLeuTrpThrLeuValSerAspPr 29
         ||||||||||||||||||||:::||||||||||||||||||||:::::
     101 TCGAACGTGCCGGCTTTCCTCAGCAAGCTGTGGACGCTTGTGGAGGAAAC 150

      29 oAspThrAspAlaLeuIleCysTrpSerProSerGlyAsnSerPheHisV 46
             |||:::   :::|||  |||||   :::|||:::|||||    |
     151 CCACACTAACGAGTTCATCACCTGGAGCCAGAATGGCCAAAGTTTTCTGG 200

      46 alPheAspGlnGlyGlnPheAlaLysGluValLeuProLysTyrPheLys 62
          ||:::|||:::   :::|||||||||||:::||||||||||||||||
     201 TCTTGGATGAGCAACGATTTGCAAAAGAAATTCTTCCCAAATATTTCAAG 250

      63 HisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheAr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CACAATAATATGGCAAGCTTTGTGAGGCAACTGAATATGTATGGTTTCCG 300

      79 gLysValValHisIleGluGlnGlyGlyLeuValLysProGluArgAsp. 95
         ||||||||||||||||:::   :::|||:::||||||  |||||||||
     301 TAAAGTAGTACATATCGAC...TCTGGAATTGTAAAGCAAGAAAGAGATG 347

      96 ..AspThrGluPheGlnHisProCysPheLeuArgGlyGlnGluGlnLeu 111
             :::|||||||||||||||  |||   :::|||||||::::::|||
     348 GTCCTGTAGAATTTCAGCATCCTTACTTCAAACAAGGACAGGATGACTTG 397

     112 LeuGluAsnIleLysArgLysValThrSerValSerThrLeuLysSerGl 128
         ||||||||||||||||||||         ||||||:::  |||  ||
     398 TTGGAGAACATTAAAAGGAAG.........GTTTCATCTTCAAAACCAGA 438

     128 uAspIleLysIleArgGlnAspSerValThrLysLeuLeuThrAspValG 145
         |:::  |||||||||||||:::::::::||||||:::::::::::::::|
     439 AGAAAATAAAATTCGTCAGGAAGATTTAACAAAAATTATAAGTAGTGCTC 488

     145 lnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMet 161
         ||   ::::::   |||||||||   ::::::|||:::|||     :::
     489 AGAAGGTTCAGATAAAACAGGAAACTATTGAGTCCAGGCTTTCTGAATTA 538

     162 LysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLy 178
         |||   |||||||||:::||||||:::||||||::::::||||||   ||
     539 AAAAGTGAGAATGAGTCCCTTTGGAAGGAGGTGTCAGAATTACGAGCAAA 588

     178 sHisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleS 195
         |||||||||||||||:::|||::::::|||::::::||||||::::::
     589 GCATGCACAACAGCAACAAGTTATTCGAAAGATTGTCCAGTTTATTGTTA 638

     195 erLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleProLeu 211
         ::||||||||:::|||:::::::::::::::::||||||||   ||||||
     639 CATTGGTTCAAAATAACCAACTTGTGAGTTTAAAACGTAAAAGGCCTCTA 688

     212 MetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSerArgGl 228
         :::||||||   :::|||:::                  ::::::::
     689 CTTCTAAACACTAATGGAGCC....................CAAAAGAA 717

     228 nPheSerLeuGluHisVal.......................HisGlyS 237
         :      ::::::|||:::                      |||   :
     718 GAACCTGTTTCAGCACATAGTCAAAGAACCAACTGATAATCATCATCATA 767

     237 erGlyProTyrSerAlaProSerProAlaTyrSerSerSerSerLeuTyr 253
         ::  |||                    :::|||   ::::::
     768 AAGTTCCA....................CACAGTAGGACTGAAGGTTTA 796

     254 AlaProAspAlaValAlaSerSerGlyProIleIleSerAspIleThr.. 269
            |||          |||:::     ||||||   |||:::|||
     797 AAGCCAAGGGAGAGGATTTCAGATGACATCATTATTTATGATGTTACTGA 846

     270 ................GluLeuAlaProAlaSerProMetAlaSerProG 281
                         |||    |||:::   |||    ::::::
     847 TGATAATGCAGATGAAGAAAATATCCCAGTTATTCCAGAAACTAATGAGG 896

     281 lyGlySerIleAspGluArgProLeuSerSerSerProLeuValArgVal 297
                   |||             |||:::   |||   :::   :::
     897 ATGTTATATCTGATCCCTCCAACTGTAGCCAGTACCCTGATATTGTCATC 946

     298 LysGluGluProProSerProProGlnSerProArgVal.........Gl 311
            |||:::          :::         :::|||   :::          ||
     947 GTTGAAGATGACAATGAAGATGAGTATGCACCTGTCATTCAGAGTGGAGA 996

     311 uGluAlaSerProGlyArgProSerSerValAspThrLeuLeuSerProT 328
         |:::   :::|||:::|||  |||
     997 GCAGAATGAACCAGCCAGAGAATCC......................... 1021

     328 hrAlaLeuIleAspSerIleLeuArgGluSerGluProAlaProAlaSer 344
                 ::::::|||             |||:::|||      :::|||
    1022 ........CTAAGTTCAGGCAGTGATGGCAGCAGCCCTCTCATGTCTAGT 1063

     345 ValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgPr 361
         ::::::   |||:::   :::    :::   |||
```

```
    1064 GCTGTCCAGCTAAATGGCTCATCCAGTCTGACC..................  1096

     361 oProSerProProProThrSerThrProGluLysCysLeu......SerV  376
         |||      |||::::::      ::::::   |||      ::::
    1097 ....TCAGAAGATCCAGTGACCATGATGGATTCCATTTTGAATGATAACA  1142

     376 alAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAsp  392
         ::      |||   |||   ||||||   |||:::||||||::::::|||
    1143 TCAATCTTTTGGGAAAGGTTGAGCTGTTGGATTATCTTGACAGTATTGAC  1192

     393 SerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerVa  409
         :::|||:::::::::|||:::||||||||||::::::   ||||||::
    1193 TGCAGTTTAGAGGACTTCCAGGCCATGCTATCAGGAAGACAATTTAGCAT  1242

     409 lAspThrSerAlaLeuLeuAspLeuPheSerProSerValThrVal...P  425
         :|||   :::   |||:::||||||||::   ||||||   :::   |
    1243 AGACCCAGATCTCCTGGTTGATCTTTTCACTAGTTCTGTGCAGATGAATC  1292

     425 roAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGlu  441
         ||        :::   :::      |||:::      :::::::::
    1293 CCACAGATTACATCAATAATACAAAATCTGAGAATAAAGGATTAGAAACT  1342

     442 LeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSerSe  458
             :::            |||         |||::::::::|||::
    1343 ACCAAGAACAATGTAGTTCAGCCAGTTTCGGAAGAGGGAAGAAAATCTAA  1392

     458 rProAspSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPheL  475
         :           |||||||||::::::|||||||||   ||||||:::
    1393 ATCCAAACCAGATAAGCAGCTTATCCAGTATACCGCCTTTCCACTTCTTG  1442

     475 euLeuAspProGly............SerValAspThrGlySerAsn...  486
           :::      |||           ||||||:::   :::|||:::
    1443 CATTCCTCGATGGGAACCCTGCTTCTTCTGTTGAACAGGCGAGTACAACA  1492

     487 ........................AspLeuProVal...LeuPheGl  493
                                 |||   |||:::   :::   ||
    1493 GCATCATCAGAAGTTTTGTCCTCTGTAGATAAACCCATAGAAGTTGATGA  1542

     493 uLeuGlyGluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAspP  510
         ||||   ::::::|||         |||                   |
    1543 GCTTCTGGATAGCAGCCTAGACCCAGAA.....................C  1571

     510 roThrIleSerLeuLeuThrGlySerGluPro 520
         |||||   |||   |||:::      ||||||
    1572 CAACCCAAAGTAAGCTTGTTCGCCTGGAGCCA 1603
```

seq_name: gb_ro:AF172640

seq_documentation_block:

```
LOCUS       AF172640     1792 bp    mRNA            ROD       25-AUG-1999
DEFINITION  Rattus norvegicus heat shock factor 2 (hsf2) mRNA, complete cds.
ACCESSION   AF172640
VERSION     AF172640.1  GI:5764552
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 1792)
  AUTHORS   Lee,S.S., Park,Y.-M. and Han,M.Y.
  TITLE     Promoter analysis of rat hsf2 gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1792)
  AUTHORS   Lee,S.S., Park,Y.-M. and Han,M.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-1999) Immune Regulation, Korea Research Institute
            of Bioscience and Biotechnology, P.O. 115, Yusung, Taejon 305-600,
            Korea
FEATURES             Location/Qualifiers
     source          1. .1792
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            1. .1792
                     /gene="hsf2"
     CDS             30. .1571
                     /gene="hsf2"
                     /note="HSF2"
                     /codon_start=1
                     /product="heat shock factor 2"
                     /protein_id="AAD51329.1"
                     /db_xref="GI:5764553"
                     /translation="MKQSSNVPAFLSKLWTLVEETHTNEFITWSQNGQSFLVLDEQRF
                     AKEILPKYFKHNNMASFVRQLNMYGFRKVVHIESGIVKQERDGPVEFQHPHFKQGQDD
                     LLENIKRKVSSSKPEENKIRQEDLTKIISSAQKVQIKQETIESRLSELKSENESLWKE
                     VSELRAKHAQQQQVIRKIVQFIVTLVQNNQLVSLKRKRPLLLNTNGAPKKNLFQHIVK
                     EPTDNHHHKVPHSRTEGLKSRERISDDIIIYDVTEDNVDEENIPVIPETNEDVVVDPS
                     SQYPDIVIVEDDNEDEYAPVIQSGEQSEPARVGSASSSSPLMSSAVQLNGSSNLTSED
                     PVTMMDSILNQNINLLGKVELLDYLDSIDCSLEDFQAMLSGRQFSIDPDLLVDSENKG
                     LEATKNSVVQQVSEEGRKSKSKPDKQLIQYTAFPLLAFLDGNPASAVEQGTTTASSEV
                     VPSVDKPIEVDELLDSNLDQEPTQSKLVRLEPLTEAEASEATLFYLCELAPAPLDSDM
                     PLLDS"
BASE COUNT      574 a    363 c    381 g    474 t
ORIGIN
```

alignment_scores:

```
            Quality:  787.00              Length:    534
              Ratio:  2.281                 Gaps:     18
 Percent Similarity:  64.607   Percent Identity: 37.640
```

alignment_block:

US-09-304-121-2 x AF172640 ..

Align seg 1/1 to: AF172640 from: 1 to: 1792

```
      13 SerAsnValProAlaPheLeuThrLysLeuTrpThrLeuValSerAspPr  29
         ||||||||||||||||||||||:::||||||||||||||||||::::::
      42 TCCAACGTGCCGGCTTTCCTCAGCAAGCTGTGGACGCTTGTAGAGGAAAC  91

      29 oAspThrAspAlaLeuIleCysTrpSerProSerGlyAsnSerPheHisV  46
             |||:::   :::|||   ||||||   :::|||:::||||||   |
      92 CCACACTAACGAGTTCATCACCTGGAGTCAGAATGGACAAAGTTTTCTGG  141

      46 alPheAspGlnGlyGlnPheAlaLysGluValLeuProLysTyrPheLys  62
         ||:::|||:::   :::||||||||||||:::|||||||||||||||||
     142 TCTTGGATGAGCAAAGATTTGCAAAGGAAATTCTCCCCAAGTACTTCAAA  191

      63 HisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheAr  79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     192 CACAATAACATGGCAAGCTTTGTGAGACAGCTAAATATGTATGGCTTCCG  241

      79 gLysValValHisIleGluGlnGlyGlyLeuValLysProGluArgAsp.  95
         ||||||||||||||||||   :::|||:::||||||   |||||||||
     242 TAAAGTAGTACATATCGAA...TCTGGAATTGTCAAACAGGAAAGAGATG  288

      96 ..AspThrGluPheGlnHisProCysPheLeuArgGlyGlnGluGlnLeu  111
            :::|||||||||||||||   |||   :::||||||::::::|||
     289 GCCCTGTTGAATTTCAGCATCCTCATTTCAAACAAGGCCAGGATGACTTG  338

     112 LeuGluAsnIleLysArgLysValThrSerValSerThrLeuLysSerGl  128
         ||||||||||||||||||||          ||||||:::   |||   ||
     339 TTGGAGAACATTAAAAGGAAG.........GTTTCGTCTTCAAAACCAGA  379

     128 uAspIleLysIleArgGlnAspSerValThrLysLeuLeuThrAspValG  145
         |:::   ||||||||||||:::::::::|||||:::::::::::::::|
     380 GGAAAATAAAATTCGTCAGGAAGATTTAACAAAAATTATTAGTAGTGCCC  429

     145 lnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMet  161
         ||   ::::::.  |||||||||   :::::::|||:::|||      :::
     430 AGAAGGTTCAAATAAAACAAGAAACTATTGAGTCCAGGCTTTCAGAATTA  479

     162 LysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLy  178
         |||   |||||||||:::|||||||:::|||||||:::::::||||||   ||
     480 AAAAGTGAGAATGAATCCCTTTGGAAGGAGGTGTCAGAGTTAAGAGCGAA  529

     178 sHisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleS  195
```

```
         ||||||||||||||:::|||::::::|||::::::||||||:::::::
 530 GCATGCACAGCAGCAGCAAGTTATTCGAAAGATTGTCCAGTTTATTGTTA 579

 195 erLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleProLeu 211
     ::|||||||||:::|||:::::::::::::::||||||||   ||||||
 580 CATTGGTTCAGAATAATCAACTTGTGAGTTTAAAACGTAAAAGGCCTCTA 629

 212 MetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSerArgGl 228
     :::||||||   :::|||:::               ||||||   :::
 630 CTTCTAAACACAAATGGAGCC............CCAAAGAAGAAT..... 662

 228 nPheSerLeuGluHisVal.......................HisGlyS 237
         ::::::|||:::                          |||    :
 663 ....CTATTTCAGCACATAGTCAAAGAACCAACCGATAATCACCATCATA 708

 237 erGlyProTyrSerAlaProSerProAlaTyrSerSerSerSerLeuTyr 253
     ::   |||:::|||        :::          |||   :::   :::
 709 AAGTTCCACACAGCAGAACTGAAGGTTTAAAGTCAAGAGAACGGATT... 755

 254 AlaProAspAlaValAlaSerSerGlyProIleIleSerAspIleThr.. 269
                          |||:::     ||||||   |||:::|||
 756 .................TCAGATGACATAATTATTTATGATGTTACTGA 787

 270 ...............GluLeuAlaProAlaSerProMetAlaSerProG 281
                    |||       |||:::   |||    ::::::
 788 GGATAATGTGGATGAAGAAAATATTCCAGTTATTCCAGAAACAAATGAGG 837

 281 lyGlySerIleAspGluArgProLeuSerSerSerProLeuValArgVal 297
         :::|||      |||   |||:::   |||     :::   :::
 838 ATGTTGTAGTTGAT......CCCTCCAGCCAGTATCCTGACATTGTCATT 881

 298 LysGluGluProProSerProProGlnSerProArgValGluGluAlaSe 314
       |||:::          :::|||   ::::::::::::::
 882 GTTGAAGATGACAATGAGGATGAGTACGCTCCTGTCATTCAGAGTGGAGA 931

 314 rProGlyArgProSerSerValAspThrLeuLeuSerProThrAlaLeuI 331
     :   ::::::|||:::   |||   :::      |||
 932 GCAGAGTGAGCCAGCCAGAGTGGGGAGTGCCAGCAGC............. 968

 331 leAspSerIleLeuArgGluSerGluProAlaProAlaSerValThrAla 347
                          |||:::|||      :::|||:::::
 969 ...................AGCAGCCCTCTCATGTCTAGTGCTGTCCAG 998

 348 LeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProProSerPr 364
     |||:::   :::   :::   |||                     |||
 999 CTAAATGGCTCCTCCAACCTGACC....................TCAGA 1027

 364 oProProThrSerThrProGluLysCysLeu......SerValAlaCysL 379
        |||:::::        :::::   |||        ::::::      |
1028 AGACCCTGTGACCATGATGGATTCCATTCTGAATGACAACATTAACCTTT 1077

 379 euAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeu 395
     ||    |||    ||||||   |||:::||||||::::::|||     :::|||
1078 TAGGAAAGGTTGAGCTGTTGGATTACCTTGACAGTATTGATTGCAGTTTA 1127

 396 AspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThrSe 412
     :::::::::|||:::|||||||||||::::::   |||||||:::|||    ::
1128 GAGGACTTCCAAGCTATGCTCTCAGGAAGACAGTTTAGCATAGACCCAGA 1177

 412 rAlaLeuLeuAspLeuPheSerProSerValThrValProAspMetSerL 429
     :    |||:::|||       :::    :::
1178 CCTTCTGGTTGATTCTGAGAATAAGGGA...................... 1205

 429 euProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLeuSerPro 445
             |||::::::::        |||:::    :::      :::|||
1206 ........TTAGAAGCTACCAAGAACAGTGTTGTTCAACAAGTTTCAGAA 1247

 446 GlnGluProProArgProProGluAlaGluAsnSerSerProAspSerGl 462
     :::                 :::    ::::::|||:::||||||
1248 GAG..............GGAAGAAAATCTAAATCCAAGCCAGAT..... 1277

 462 yLysGlnLeuValHisTyrThrAlaGlnProLeu......PheLeu.... 475
     *|||||||||::::::||||||||     ||||||         ||||||

1278 .AAACAGCTTATCCAGTACACTGCCTTTCCACTTCTCGCATTCCTTGATG 1326

 476 ................LeuAspProGlySerValAspThrGlySerAsn 486
                    ::::::   |||::::::    ::::::|||:::
1327 GGAACCCTGCATCTGCTGTTGAACAGGGGACTACAACAGCATCGTCAGAA 1376

 487 AspLeuProValLeuPheGluLeuGlyGluGlySerTyrPheSerGluGl 503
        :::|||   :::    :::        |||    :::   :::   :::::
1377 GTTGTACCCTCTGTAGATAAGCCTATAGAAGTTGATGAGCTCCTGGATAG 1426

 503 yAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGlySerGluP 520
     ::::   :::   ::::::||||||   |||   |||:::       ||||
1427 CAAC...CTAGACCAAGAACCAACCCAAAGTAAGCTTGTCCGCCTGGAAC 1473

 520 ro 520
     ||
1474 CA 1475
```

seq_name: gb_ro:MMHSF2

```
seq_documentation_block:
LOCUS       MMHSF2       1972 bp    mRNA            ROD       28-NOV-1991
DEFINITION  M.musculus mRNA for heat shock transcripton factor 2.
ACCESSION   X61754
VERSION     X61754.1  GI:51447
KEYWORDS    heat shock transcription factor; HSF2.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1972)
  AUTHORS   Sarge,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1991) K.D. Sarge, Northwestern University, Dep of
            Biochem.,Mol. and Cell. Biology, Hogan 2-100, 2153 Sheridan Rd.,
            Evanston IL 60208, USA
REFERENCE   2  (bases 1 to 1972)
  AUTHORS   Sarge,K.D., Zimarino,V., Holm,K., Wu,C. and Morimoto,R.I.
  TITLE     Cloning and characterization of two mouse heat shock factors with
            distinct inducible and constitutive DNA-binding ability
  JOURNAL   Genes Dev. 5 (10), 1902-1911 (1991)
  MEDLINE   92009180
COMMENT     See also x61753.
FEATURES             Location/Qualifiers
     source          1. .1972
                     /organism="Mus musculus"
                     /strain="Balb/C"
                     /db_xref="taxon:10090"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="Clontech lambda gt11 liver cDNA"
                     /clone="C9"
     CDS             159. .1712
                     /codon_start=1
                     /product="heat shock transcription factor 2"
                     /protein_id="CAA43893.1"
                     /db_xref="GI:51448"
                     /db_xref="MGD:MGI:96239"
                     /db_xref="SWISS-PROT:P38533"
                     /translation="MKQSSNVPAFLSKLWTLVEETHTNEFITWSQNGQSFLVLDEQRF
                     AKEILPKYFKHNNMASFVRQLNMYGFRKVVHIESGIIKQERDGPVEFQHPYFKQGQDD
                     LLENIKRKVSSSKPEENKIRQEDLTKIISSAQKVQIKQETIESRLSELKSENESLWKE
                     VSELRAKHAQQQQVIRKIVQFIVTLVQNNQLVSLKRKRPLLLNTNGAPKKNLYQHIVK
                     EPTDNHHHKVPHSRTEGLKSRERISDDIIIYDVTDDNVDEENIPVIPETNEDVVVDSS
                     NQYPDIVIVEDDNEDEYAPVIQSGEQSEPAREPLRVGSAGSSSPLMSSAVQLNGSSSL
                     TSEDPVTMMDSILNDNINLLGKVELLDYLDSIDCSLEDFQAMLSGRQFSIDPDLLVDS
                     ENKGLEATKSSVVQHVSEEGRKSKSKPDKQLIQYTAFPLLAFLDGNSASAIEQGSTTA
                     SSEVVPSVDKPIEVDELLDSSLDPEPTQSKLVRLEPLTEAEASEATLFYLCELAPAPL
                     DSDMPLLDS"
BASE COUNT      604 a    420 c    432 g    516 t
ORIGIN
```

```
alignment_scores:
           Quality:  774.00              Length:    544
             Ratio:  2.263                Gaps:     22
 Percent Similarity:  62.868   Percent Identity: 37.868

alignment_block:
 US-09-304-121-2 x MMHSF2   ..

 Align seg 1/1  to: MMHSF2  from: 1  to: 1972

      13 SerAsnValProAlaPheLeuThrLysLeuTrpThrLeuValSerAspPr 29
     171 TCCAACGTGCCGGCTTTCCTCAGCAAGCTGTGGACGCTTGTGGAGGAAAC 220

      29 oAspThrAspAlaLeuIleCysTrpSerProSerGlyAsnSerPheHisV 46
     221 CCACACCAACGAGTTCATCACCTGGAGTCAGAATGGACAAAGTTTTCTGG 270

      46 alPheAspGlnGlyGlnPheAlaLysGluValLeuProLysTyrPheLys 62
     271 TCTTGGATGAGCAAAGATTTGCAAAGGAAATTCTTCCTAAGTACTTCAAA 320

      63 HisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheAr 79
     321 CACAATAACATGGCGAGCTTTGTGAGACAACTAAATATGTATGGCTTCCG 370

      79 gLysValValHisIleGluGlnGlyGlyLeuValLysProGluArgAsp. 95
     371 AAAAGTAGTGCATATCGAA...TCTGGAATTATCAAACAGGAAAGAGATG 417

      96 ..AspThrGluPheGlnHisProCysPheLeuArgGlyGlnGluGlnLeu 111
     418 GCCCTGTTGAATTTCAGCATCCTTATTTCAAGCAAGGCCAGGATGACCTG 467

     112 LeuGluAsnIleLysArgLysValThrSerValSerThrLeuLysSerGl 128
     468 TTGGAGAACATTAAAAGGAAG.........GTTTCATCTTCAAAACCAGA 508

     128 uAspIleLysIleArgGlnAspSerValThrLysLeuLeuThrAspValG 145
     509 GGAAAATAAAATTCGTCAGGAAGATTTAACAAAAATTATTAGTAGTGCTC 558

     145 lnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMet 161
     559 AGAAGGTTCAAATAAAACAAGAAACTATTGAGTCCAGGCTTTCAGAATTA 608

     162 LysHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLy 178
     609 AAAAGTGAGAATGAATCCCTTTGGAAGGAGGTGTCAGAACTAAGAGCAAA 658

     178 sHisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleS 195
     659 GCATGCCCAGCAGCAACAAGTTATTCGGAAGATTGTCCAGTTTATTGTTA 708

     195 erLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleProLeu 211
     709 CATTGGTTCAGAATAATCAACTTGTGAGTTTAAAACGTAAAAGGCCTCTA 758

     212 MetLeuAsnAspSerGlySerAlaHisSerMetProLysTyrSerArgGl 228
     759 CTTCTAAACACAAATGGAGCC.............CCAAAGAAGAAT..... 791

     228 nPheSerLeuGluHisVal.......................HisGlyS 237
     792 ....CTATATCAGCACATAGTCAAAGAACCAACTGATAATCACCATCATA 837

     237 erGlyProTyrSerAlaProSerProAlaTyrSerSerSerSerLeuTyr 253
     838 AAGTTCCACACAGCAGGACTGAAGGTTTAAAGTCAAGAGAACGGATT... 884

     254 AlaProAspAlaValAlaSerSerGlyProIleIleSerAspIleThr.. 269
     885 ..................TCAGATGACATAATTATTTATGATGTTACTGA 916

     270 ................GluLeuAlaProAlaSerProMetAlaSerProG 281
     917 CGATAATGTGGATGAAGAAAATATTCCAGTTATTCCAGAAACAAATGAGG 966

     281 lyGlySerIleAspGluArgProLeuSerSerSerProLeuValArgVal 297
     967 ATGTTGTAGTGGATTCCTCC......AACCAGTATCCTGACATTGTCATT 1010

     298 LysGluGluProProSerProProGlnSerProArgVal.........Gl 311
    1011 GTTGAAGATGACAACGAGGATGAGTATGCTCCTGTCATTCAGAGTGGAGA 1060

     311 uGluAlaSerProGlyArg...ProSerSerValAspThrLeuLeuSerP 327
    1061 GCAGAGTGAACCAGCCAGAGAACCCTTACGTGTGGGGAGTGCTGGCAGC. 1109

     327 roThrAlaLeuIleAspSerIleLeuArgGluSerGluProAlaProAla 343
    1110 ..............................AGCAGCCCTCTCATGTCT 1127

     344 SerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyAr 360
    1128 AGTGCTGTCCAGCTAAACGGCTCCTCCAGTCTGACC.............. 1163

     360 gProProSerProProProThrSerThrProGluLysCysLeu......S 375
    1164 .......TCAGAAGACCCTGTGACCATGATGGACTCCATTCTGAATGACA 1206

     375 erValAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMet 391
    1207 ACATTAACCTGTTAGGAAAGGTTGAGCTGTTGGATTACCTTGACAGTATT 1256

     392 AspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSe 408
    1257 GATTGCAGTTTAGAGGACTTCCAAGCTATGCTCTCAGGAAGACAGTTTAG 1306

     408 rValAspThrSerAlaLeuLeuAspLeuPheSerProSerValThrValP 425
    1307 CATAGACCCAGATCTTCTGGTTGATTCTGAGAATAAGGGA.......... 1346

     425 roAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGlu 441
    1347 ...................CTAGAAGCTACCAAGAGCAGTGTTGTTCAA 1376

     442 LeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSerSe 458
    1377 CATGTGTCAGAAGAG..............GGAAGAAAATCTAAATCCAA 1411

     458 rProAspSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeu.... 473
    1412 GCCAGAC......AAACAACTTATCCAGTATACTGCCTTTCCACTTCTTG 1455

     474 ..PheLeu.....................LeuAspProGlySerValAsp 482
    1456 CATTCCTGGATGGGAACTCTGCATCTGCTATTGAACAGGGGAGTACAACT 1505

     483 ThrGlySerAsn..........*.....AspLeuProVal...LeuPheGl 493
    1506 GCATCGTCAGAAGTTGTGCCTTCTGTAGATAAACCCATAGAAGTCGATGA 1555

     493 uLeuGlyGluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAspP 510
    1556 GCTCCTGGATAGCAGCCTGGATCCAGAA.....................C 1584

     510 roThrIleSerLeuLeuThrGlySerGluPro 520
    1585 CGACCCAGAGTAAGCTTGTCCGCCTGGAACCA 1616

seq_name: gb_pr2:AB029348

seq_documentation_block:
LOCUS       AB029348     1577 bp    mRNA            PRI       23-SEP-1999
```

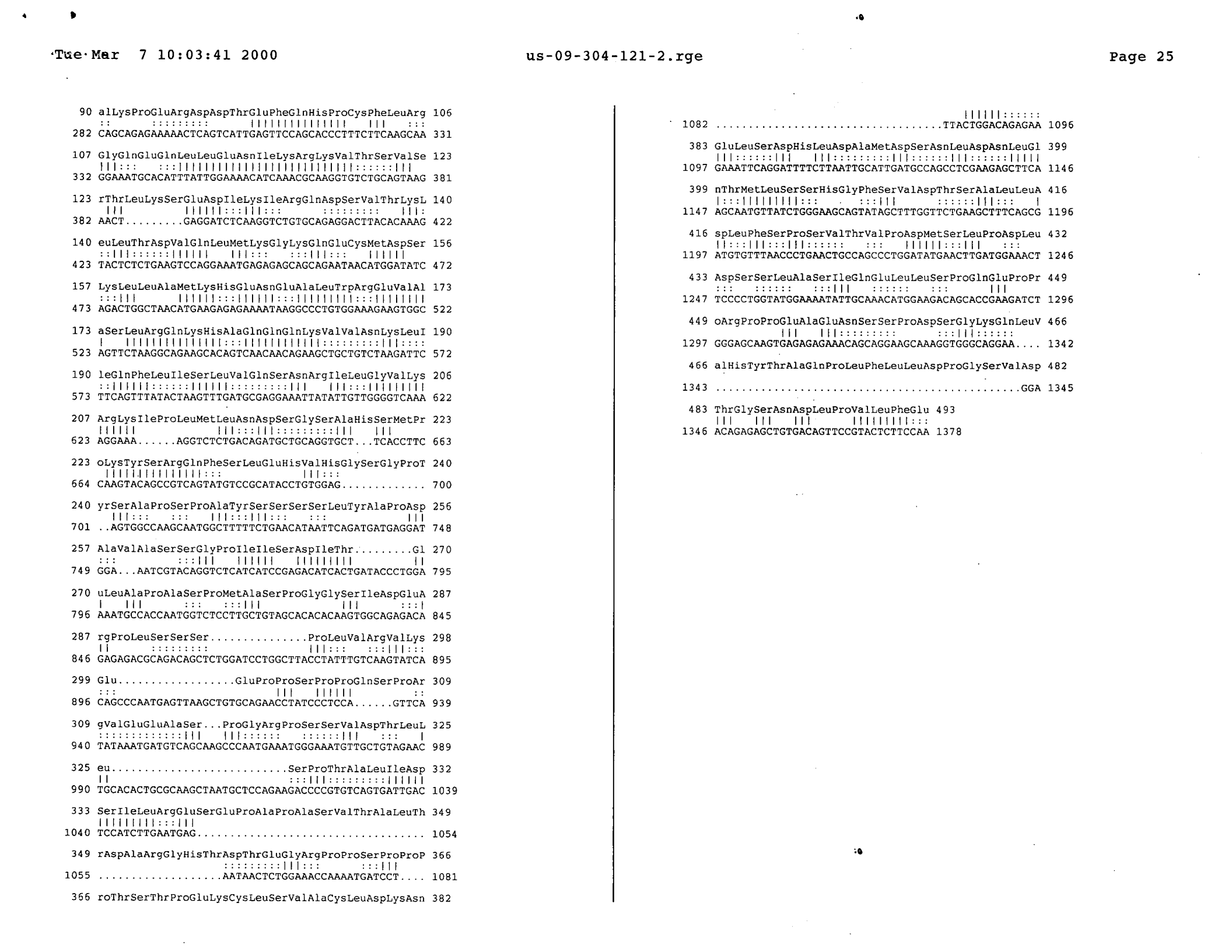
```
DEFINITION  Homo sapiens mRNA for transcription factor HSF4b isoform, complete
            cds.
ACCESSION   AB029348
VERSION     AB029348.1  GI:5921134
KEYWORDS    transcription factor HSF4b isoform.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Tanabe,M., Sasai,N., Nagata,K., Liu,X.D., Liu,P.C., Thiele,D.J. and
            Nakai,A.
  TITLE     The mammalian HSF4 gene generates both an activator and a repressor
            of heat shock genes by alternative splicing
  JOURNAL   J. Biol. Chem. 274 (39), 27845-27856 (1999)
  MEDLINE   99419073
REFERENCE   2  (bases 1 to 1577)
  AUTHORS   Nakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) to the DDBJ/EMBL/GenBank databases. Akira
            Nakai, Institute for Frontier Medical Sciences, Department of
            Molecular and Cell Biology; Sakyo-ku, Kyoto 606-8397, Japan
            (E-mail:nakai@frontier.kyoto-u.ac.jp, Tel:81-75-751-4638,
            Fax:81-75-752-9017)
FEATURES             Location/Qualifiers
     source          1..1577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             5..1486
                     /codon_start=1
                     /product="transcription factor HSF4b isoform"
                     /protein_id="BAA84582.1"
                     /db_xref="GI:5921135"
                     /translation="MVQEAPAALPTEPGPSPVPAFLGKLWALVGDPGTDHLIRWSPSG
                     TSFLVSDQSRFAKEVLPQYFKHSNMASFVRQLNMYGFRKVVSIEQGGLLRPERDHVEF
                     QHPSFVRGREQLLERVRRKVPALRGDDGRWRPEDLGRLLGEVQALRGVQESTEARLRE
                     LRQQNEILWREVVTLRQSHGQQHRVIGKLIQCLFGPLQAGPSNAGGKRKLSLMLDEGS
                     SCPTPAKFNTCPLPGALLQDPYFIQSPLPETNLGLSPHRARGPIISDIPEDSPSPEGT
                     RLSPSSDGRREKGLALLKEEPASPGGDGEAGLALAPNECDFCVTAPPPLPVAVVQAIL
                     EGKGSFSPEGPRNAQQPEPGDPREIPDRGPLGLESGDRSPESLLPPMLLQPPQESVEP
                     AGPLDVLGPSLQGREWTLMDLDMELSLMQPLVPERGEPELAVKGLNSPSPGKDPTLGA
                     PLLLDVQAALGGPALGLPGALTIYSTPESRTASYLGPEASPSP"
BASE COUNT      300 a    494 c    523 g    260 t
ORIGIN


alignment_scores:
          Quality:  763.00            Length:     534
            Ratio:  2.312               Gaps:      18
 Percent Similarity:  61.798   Percent Identity: 40.075

alignment_block:
 US-09-304-121-2 x AB029348   ..

 Align seg 1/1  to: AB029348  from: 1  to: 1577

      3 LeuProValGlyProGlyAlaAlaGlyProSerAsnValProAlaPheLe 19
        ||||||:::   |||         |||||||||   ||||||||||||||
     29 CTGCCCACGGAGCCA.........GGCCCCAGCCCCGTGCCTGCCTTCCT 69

     19 uThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaLeuIleC 36
        |   |||||||||:::||||||:::||||||   ||||||   ||||||
     70 CGGCAAGCTATGGGCGCTGGTGGGGGACCCAGGCACAGACCACCTGATCC 119

     36 ysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGlyGlnPhe 52
          |||||||||||||||:::||||||   |||   ||||||::::::|||
    120 GCTGGAGCCCGAGCGGGACCAGTTTCCTCGTAAGCGACCAGAGCCGTTTC 169

     53 AlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAlaSerPh 69
        ||||||||||||||||||:::||||||||||||:::||||||||||||||
    170 GCCAAGGAAGTGCTGCCCCAGTATTTCAAGCATAGCAACATGGCGAGCTT 219

     69 eValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisIleGluG 86
        ||||||||||||||||||||||||||||||||||||||||   |||||||
    220 CGTGCGCCAACTCAACATGTACGGTTTTCGGAAGGTGGTGAGCATCGAGC 269

     86 lnGlyGlyLeuValLysProGluArgAspAspThrGluPheGlnHisPro 102
        ||||||||||::::::||||||||||||   :::||||||||||||||
    270 AGGGCGGCCTGCTTAGGCCGGAGCGCGACCACGTCGAGTTCCAGCACCCG 319

    103 CysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysVa 119
           |||:::||||||:::|||||||||||||||:::::::::||||||||
    320 AGCTTCGTGCGCGGCCGCGAGCAGCTACTGGAGCGCGTGCGGCGCAAGGT 369

    119 lThrSerValSerThrLeuLysSerGluAspIleLysIleArgGlnAspS 136
        |   :::         |||:::::::::|||   :::   |||   ::::
    370 GCCCGCG.........CTGCGCGGCGACGACGGCCGCTGGCGCCCGGAGG 410

    136 erValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGlu 152
        :::::   :::||||||   :::||||||   ::::::|||   ||||||
    411 ACCTGGGTCGACTACTGGGCGAGGTGCAGGCTTTGCGGGGAGTGCAGGAG 460

    153 CysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTr 169
              :::::::::|||      ::::::::::::||||||   |||||
    461 AGCACCGAGGCGCGGCTGCGGGAGCTCAGGCAGCAGAACGAGATCTTGTG 510

    169 pArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValV 186
        |||||||||::::::|||||||||:::|||:::|||||||:::::::|||:
    511 GCGGGAGGTGGTGACACTTCGGCAGAGCCACGGTCAGCAGCACCGGGTCA 560

    186 alAsnLysLeuIleGlnPheLeuIleSerLeuValGln.........Ser 199
        :::::||||||||||||   |||::::::   :::|||         |||
    561 TTGGCAAGCTGATCCAGTGTCTCTTTGGGCCACTTCAGGCGGGGCCGAGC 610

    200 AsnArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSe 216
        |||         |||   ||||||||||:::   |||||||||   |||::
    611 AAT......GCAGGAGGCAAGAGAAAGCTGTCCCTGATGCTG...GATGA 651

    216 rGlySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluH 233
        :||||||:::            |||         :::         |||
    652 GGGGAGCTCATGCCCAACACCTGCCAAGTTCAACACCTGCCCTCTACCTG 701

    233 isValHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSer 249
          :::      :::   ||||||         ||||||      ::::::
    702 GTGCCCTTCTGCAGGACCCCTACTTCATCCAGTCGCCTCTCCCAGAGACA 751

    250 SerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSe 266
        :::            :::|||            :::   |||||||||||||
    752 AATTTGGGCCTTAGCCCTCAC......AGGGCCAGGGGCCCCATCATCTC 795

    266 rAspIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlyS 283
        |||||||   |||   :::|||:::               :::   :::   |
    796 TGACATCCCAGAAGACTCTCCATCCCCTGAGGGGACCAGGCTTTCTCCCT 845

    283 erIleAspGluArgProLeuSerSerSerProLeuValArgValLysGlu 299
        ||   |||   |||         ::::::   |||:::   :::||||||
    846 CCAGTGATGGCAGG......AGGGAGAAGGGCCTGGCACTGCTCAAAGAA 889

    300 GluProProSerProProGlnSerProArgValGluGluAlaSerProGl 316
        ||||||                              ||||||||||
    890 GAGCCG..............................GCCAGTCCAGG 906

    316 yArgProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspS 333
        |      :::::::::         |||:::|||:::         |||
    907 GGGGGATGGCGAGGCCGGGCTGGCCCTGGCCCCAAACGAG...TGTGACT 953

    333 erIleLeuArgGluSerGluProAlaPro...AlaSerValThrAlaLeu 348
          :::               |||   |||   |||   |||   |||:::
    954 TCTGCGTGACAGCCCCCCCGCCACTGCCTGTGGCTGTGGTGCAGGCCATC 1003

    349 ThrAspAlaArgGlyHisThrAspThrGluGlyArgProProSerProPr 365
          :::::::::|||      :::   ||||||            |||
   1004 CTGGAAGGGAAAGGGAGCTTCAGCCCCGAGGGG...........CCCAG 1041

    365 oProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysA 382
           ::::::   ||||||                           |||   :
```

```
  1042 GAATGCCCAACAGCCTGAACCAGGG.................GATCCCA 1073

   382 snGluLeuSerAspHis............LeuAspAlaMetAspSerAsn 394
       ::|||:::   |||:::           |||::::::   |||   :::
  1074 GGGAGATACCTGACAGGGGGCCTCTGGGCCTGGAAAGCGGGGACAGGAGC 1123

   395 LeuAspAsnLeu......GlnThrMetLeuSerSerHisGlyPheServa 409
          :::::::|||            :::|||:::            |||||
  1124 CCAGAGAGTCTGCTGCCTCCGATGCTGCTTCAGCCCCCTCAAGAAAGTGT 1173

   409 lAspThrSerAlaLeuLeuAspLeuPheSerProSerValThrValProA 426
       |:::   ::::::   ||||||:::::::::||||||:::         :
  1174 GGAACCTGCAGGGCCTCTAGATGTGCTGGGCCCCAGTCTCCAAGGGCGAG 1223

   426 spMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeu 442
       ::   :::|||   |||||||||   :::|||:::   :::|||   |||
  1224 AATGGACCCTGATGGACTTGGACATGGAGCTGTCCTTGATGCAGCCCTTG 1273

   443 LeuSerProGlnGluProProArgProProGluAlaGluAsnSerSerPr 459
       :::      :::      |||:::                  :::|||||
  1274 GTTCCAGAGCGGGGTGAGCCTGAGCTGGCGGTCAAGGGGTTAAATTCTCC 1323

   459 oAspSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuL 476
       |:::   ||||||:::   :::         :::   ||||||   ||||
  1324 AAGCCCAGGGAAGGACCCCACG......CTCGGGGCCCCACTC...CTGC 1364

   476 euAspProGlySerValAspThrGlySerAsnAspLeuProValLeuPhe 492
       |||||      ::::::      |||            |||:::|||
  1365 TGGATGTCCAGGCGGCCTTGGGAGGC............CCAGCCCTGGGC 1402

   493 GluLeuGlyGluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAs 509
             |||      :::   :::|||                  |||::
  1403 CTGCCTGGGGCTTTAACCATTTATAGCACT............CCTGAGAG 1440

   509 pProThrIleSerLeuLeuThrGlySerGluProProLysAlaLysAspP 526
       :   |||   |||   |||      :::      |||:::      ||||
  1441 CCGGACTGCCTCCTACTTGGGCCCGGAAGCCAGTCCCTCCCCCTAAGACC 1490

   526 ro 526
       ||
  1491 CC 1492

seq_name: gb_ro:AB029349

seq_documentation_block:
LOCUS       AB029349     1686 bp    mRNA            ROD       23-SEP-1999
DEFINITION  Mus musculus mRNA for transcription factor HSF4b isoform, complete
            cds.
ACCESSION   AB029349
VERSION     AB029349.1  GI:5921136
KEYWORDS    transcription factor HSF4b isoform.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Tanabe,M., Sasai,N., Nagata,K., Liu,X.D., Liu,P.C., Thiele,D.J. and
            Nakai,A.
  TITLE     The mammalian HSF4 gene generates both an activator and a repressor
            of heat shock genes by alternative splicing
  JOURNAL   J. Biol. Chem. 274 (39), 27845-27856 (1999)
  MEDLINE   99419073
REFERENCE   2  (bases 1 to 1686)
  AUTHORS   Nakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) to the DDBJ/EMBL/GenBank databases. Akira
            Nakai, Institute for Frontier Medical Sciences, Department of
            Molecular and Cell Biology; Sakyo-ku, Kyoto 606-8397, Japan
            (E-mail:nakai@frontier.kyoto-u.ac.jp, Tel:81-75-751-4638,
            Fax:81-75-752-9017)
FEATURES             Location/Qualifiers
     source          1. .1686
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             6. .1484
                     /codon_start=1
                     /product="transcription factor HSF4b isoform"
                     /protein_id="BAA84583.1"
                     /db_xref="GI:5921137"
                     /translation="MQEAPAALPTEPGPSPVPAFLGKLWALVGDPGTDHLIRWSPSGT
                     SFLVSDQSRFAKEVLPQYFKHSNMASFVRQLNMYGFRKVVSIEQGGLLRPERDHVEFQ
                     HPSFVRGREQLLERVRRKVPALRGDDSRWRPEDLSRLLGEVQALRGVQESTEARLQEL
                     RQQNEILWREVVTLRQSHSQQHRVIGKLIQCLFGPLQTGPSSTGAKRKLSLMLDEGSA
                     CSASAKFNACPVSGALLQDPYFIQSPLPETTLGLSPHRARGPIISDIPEDSPSPEGHR
                     LSPSGGCRRVKGLALLKEEPASPGGDGEAGLALAPNECDFCVTAPPPLPVAVVQAILE
                     GKGSYSPEGPRSVQQPEPRGPREVPDRGTLGLDRGNRSPESLLPPMLLRPAPETLEPV
                     APVDVLGPSLHGREWTLMDLDMELSLMQPLAPETDEAELTVKELNSSGVGKDHTLGTP
                     LMLDVQADLEGAALSVPGALTLYNVTESNASYLDPGASPSSP"
BASE COUNT      366 a    521 c    500 g    299 t
ORIGIN


alignment_scores:
            Quality:  762.00              Length:    525
              Ratio:  2.275                 Gaps:     23
 Percent Similarity:  63.810   Percent Identity: 40.381

alignment_block:
 US-09-304-121-2 x AB029349

 Align seg 1/1  to: AB029349  from: 1  to: 1686

       3 LeuProValGlyProGlyAlaAlaGlyProSerAsnValProAlaPheLe 19
         ||||||:::   |||        |||||||||   ||||||||||||||
      27 CTGCCCACGGAGCCA.........GGCCCCAGCCCGGTACCTGCCTTCCT 67

      19 uThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaLeuIleC 36
         |   |||||||||:::||||||:::|||||||   ||||||   ||||||
      68 CGGCAAGCTATGGGCGCTGGTAGGCGACCCAGGCACCGACCACCTCATCC 117

      36 ysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGlyGlnPhe 52
         ||||||||||||||:::||||||   |||   ||||||::::::|||
     118 GCTGGAGCCCGAGTGGCACCAGCTTCCTCGTAAGTGATCAGAGCCGCTTC 167

      53 AlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAlaSerPh 69
         ||||||||||||||||||:::||||||||||||:::|||||||||||||
     168 GCCAAGGAAGTGCTGCCCCAGTACTTCAAGCACAGCAACATGGCGAGCTT 217

      69 eValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisIleGluG 86
         |||||||||||||||||||||||||||||||||||||||   |||||||
     218 TGTTCGCCAACTCAACATGTATGGTTTTCGGAAGGTGGTGAGCATCGAGC 267

      86 lnGlyGlyLeuValLysProGluArgAspAspThrGluPheGlnHisPro 102
         ||||||||||:::::|||||||||||   :::|||||||||||||||
     268 AAGGTGGCCTGCTCAGACCAGAGCGTGACCACGTTGAGTTTCAGCATCCG 317

     103 CysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysVa 119
            |||:::||||||:::|||||||||||||||:::::::::|||||||
     318 AGCTTCGTGCGAGGCCGCGAGCAGCTACTGGAGCGCGTTCGCCGCAAGGT 367

     119 lThrSerValSerThrLeuLysSerGluAspIleLysIleArgGlnAspS 136
         |   :::         |||::::::::|||   :::   |||   ::::
     368 ACCTGCG.........CTGCGAGGCGATGACAGTCGATGGCGTCCGGAAG 408

     136 erValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGlu 152
         ::::::::::|||||   :::||||||   ::::::|||   ||||||
     409 ACCTGAGCCGACTGCTGGGAGAGGTGCAAGCTTTGAGAGGAGTGCAGGAG 458

     153 CysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTr 169
               ::::::::|||      ::::::::::||||||   |||||
     459 AGCACGGAGGCACGGCTGCAGGAACTCAGGCAGCAGAACGAGATCTTGTG 508

     169 pArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValV 186
         ||||||||||:::::|||||||||:::|||:::|||||:::::|||:
     509 GCGAGAGGTGGTAACTTTGCGGCAGAGCCACAGTCAGCAGCACCGGGTCA 558
```

```
 186 alAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn...Arg 201
     :::::|||||||||||||   |||::::::   :::|||::::::
 559 TCGGCAAGCTAATCCAGTGCCTGTTTGGGCCACTTCAGACAGGGCCCAGC 608

 202 IleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySe 218
         |||:::|||||||||:::   |||||||||   |||:::|||||
 609 AGTACAGGAGCCAAGAGAAAACTGTCCCTAATGCTA...GATGAGGGGAG 655

 218 rAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisValH 235
     ||||   |||           |||   :::|||:::        |||
 656 CGCCTGTTCAGCA.........TCTGCCAAATTCAATGCCTGCCCTGTGT 696

 235 isGlySer.........GlyProTyrSerAlaProSerProAlaTyrSer 248
       |||:::          ||||||         ||||||      :::
 697 CTGGTGCCCTCCTCCAGGACCCCTACTTTATCCAGTCGCCCCTCCCAGAG 746

 249 SerSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIl 265
     ::::::         :::|||            :::   ||||||||||
 747 ACCACGTTGGGCCTCAGCCCTCAC......AGGGCCAGAGGGCCCATCAT 790

 265 eSerAspIleThrGluLeuAlaProAlaSerPro.........MetAlaS 279
     ||||||||||   |||   :::|||   ||||||          |
 791 CTCTGACATCCCAGAAGATTCTCCA...TCTCCTGAAGGACACAGGCTTT 837

 279 erProGlyGlySerIleAspGluArgProLeuSerSerSerProLeuVal 295
     |||||:::|||:::      :::                :::   |||:::
 838 CTCCCTCCGGTGGTTGCAGGAGG............GTGAAGGGCCTGGCA 875

 296 ArgValLysGluGluProProSerProProGlnSerProArgValGluGl 312
       :::||||||||||||
 876 CTGCTCAAAGAAGAGCCG.............................. 893

 312 uAlaSerProGlyArgProSerSerValAspThrLeuLeuSerProThrA 329
      ||||||||||||     :::::::::        |||:::|||:::
 894 .GCCAGTCCAGGGGGGGATGGCGAGGCCGGGCTGGCCCTGGCCCCAAACG 942

 329 laLeuIleAspSerIleLeuArgGluSerGluProAlaPro...AlaSer 344
           |||      ::            |||   |||   |||
 943 AG...TGTGACTTCTGCGTGACAGCACCCCCACCGCTGCCCGTGGCTGTG 989

 345 ValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGly..... 359
     |||   |||:::   ::::::::::|||      :::   ||||||
 990 GTGCAGGCCATCCTGGAAGGGAAAGGGAGCTACAGCCCTGAGGGGCCCAG 1039

 360 ..........ArgProProSerProProProThrSerThrProGluLysC 373
               :::|||            |||   ::::::|||::::::
1040 GAGTGTACAACAGCCTGAACCAAGGGGCCCCAGGGAGGTACCTGACAGG. 1088

 373 ysLeuSerValAlaCysLeuAspLysAsnGluLeuSerAspHisLeuAsp 389
       ::::::      ||||||:::::::::   |||
1089 .....GGAACTCTGGGCCTGGATCGGGGTAACCGGAGCCCA......... 1124

 390 AlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGl 406
       :::|||   |||      :::      :::|||
1125 ......GAGAGTCTGCTACCCCCAATG......CTGCTTCGGCCTGCCCC 1162

 406 yPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerValT 423
       :::::::::      |||   :::|||::::::::::||||||:::
1163 TGAAACTCTGGAGCCCGTGGCACCTGTGGATGTGCTCGGCCCTAGCCTGC 1212

 423 hrValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIle 439
           :::   :::|||   |||||||||   :::|||:::   :::
1213 ATGGACGAGAATGGACCTTGATGGATCTGGACATGGAGTTGTCTCTGATG 1262

 440 GlnGluLeuLeuSerProGlnGluProProArgProProGluAlaGlu.. 455
     |||   |||                  |||:::      |||||||||
1263 CAGCCCTTG..............GCACCAGAGACGGATGAAGCAGAGCT 1297

 456 ...............AsnSerSerProAspSerGlyLysGlnLeuValH 467
                    |||||||||       |||||||:::      |
1298 GACGGTCAAGGAGTTGAATTCTTCT...GGTGTAGGGAAAGAC......C 1338

 467 isTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThr 483
```

```
      ||      :::   ||||||:::|||         ::::::::::::
1339 ACACACTGGGAACTCCACTCATGCTAGATGTCCAGGCGGATTTGGAGGGT 1388

 484 GlySerAsnAspLeuPro.........ValLeuPheGluLeuGlyGlu.. 496
     ::::::   ::::::|||         :::|||:::::::::   |||
1389 GCAGCCCTGTCTGTGCCTGGGGCTTTAACCCTGTACAACGTCACCGAGAG 1438

 497 ....GlySerTyrPheSerGluGly 503
         :::||||||::::::   |||
1439 CAACGCCTCCTACTTGGATCCTGGA 1463
```

```
seq_name: gb_ov:CHKHSF3C

seq_documentation_block:
LOCUS       CHKHSF3C     2675 bp    mRNA            VRT       17-JUN-1993
DEFINITION  Chicken mRNA sequence.
ACCESSION   L06126
VERSION     L06126.1  GI:289816
KEYWORDS    .
SOURCE      Gallus gallus (sub_species domesticus) (library: lambda ZAPII)
            embryo blood cDNA to mRNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2675)
  AUTHORS   Nakai,A. and Morimoto,R.I.
  TITLE     Characterization of a novel chicken heat sock transcription factor,
            heat shock factor 3, suggests a new regulatory pathway
  JOURNAL   Mol. Cell. Biol. 13, 1983-1997 (1993)
  MEDLINE   93204945
FEATURES             Location/Qualifiers
     source          1. .2675
                     /organism="Gallus gallus"
                     /sub_species="domesticus"
                     /db_xref="taxon:9031"
                     /cell_type="red blood cell"
                     /dev_stage="embryo"
                     /tissue_type="blood"
                     /tissue_lib="lambda ZAPII"
BASE COUNT      731 a    587 c    651 g    706 t
ORIGIN


alignment_scores:
            Quality:  740.00            Length:    511
              Ratio:  2.270               Gaps:     14
 Percent Similarity:  63.796   Percent Identity: 36.595

alignment_block:
 US-09-304-121-2 x CHKHSF3C   ..

 Align seg 1/1  to: CHKHSF3C  from: 1  to: 2675

       7 ProGlyAlaAlaGlyProSerAsnValProAlaPheLeuThrLysLeuTr 23
         |||||||||   |||   :::   ||||||:::||||||:::||||||||
      32 CCCGGCGCCCCCGGCGCCGCCCCGGTGCCTGGCTTCCTGGCCAAGCTGTG 81

      23 pThrLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerProS 40
         |:::|||||||:::|||||||::::::|||   :::||||||||||||   :
      82 GGCCCTGGTGGAGGATCCGCAGAGCGACGACGTCATCTGCTGGAGCCGGA 131

      40 erGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGluVal 56
         ::|||:::::|||   :::::::|||:::   :::||||||||||||:::
     132 ATGGCGAGAACTTCTGCATCCTGGATGAGCAGCGGTTTGCCAAGGAGCTG 181

      57 LeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGlnLe 73
         ||||||||||||||||||||||||||:::::|||||||:::|||||||||
     182 CTCCCCAAGTACTTCAAACACAACAACATCTCCAGCTTCATCCGACAGCT 231

      73 uAsnMetTyrGlyPheArgLysValValHisIleGluGlnGlyGlyLeuV 90
         |||||||||||||||||||||||||||||   :::|||:::|||   ::::
     232 CAACATGTATGGTTTCAGGAAGGTGGTTGCTTTGGAGAATGGTATGATTA 281
```


90 alLysProGluArgAspAspThrGluPheGlnHisProCysPheLeuArg 106
282 CAGCAGAGAAAAACTCAGTCATTGAGTTCCAGCACCCTTTCTTCAAGCAA 331
107 GlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerValSe 123
332 GGAAATGCACATTTATTGGAAAACATCAAACGCAAGGTGTCTGCAGTAAG 381
123 rThrLeuLysSerGluAspIleLysIleArgGlnAspSerValThrLysL 140
382 AACT.....GAGGATCTCAAGGTCTGTGCAGAGGACTTACACAAAG 422
140 euLeuThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAspSer 156
423 TACTCTCTGAAGTCCAGGAAATGAGAGAGCAGCAGAATAACATGGATATC 472
157 LysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluValAl 173
473 AGACTGGCTAACATGAAGAGAGAAAATAAGGCCCTGTGGAAAGAAGTGGC 522
173 aSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLysLeuI 190
523 AGTTCTAAGGCAGAAGCACAGTCAACAACAGAAGCTGCTGTCTAAGATTC 572
190 leGlnPheLeuIleSerLeuValGlnSerAsnArgIleLeuGlyValLys 206
573 TTCAGTTTATACTAAGTTTGATGCGAGGAAATTATATTGTTGGGGTCAAA 622
207 ArgLysIleProLeuMetLeuAsnAspSerGlySerAlaHisSerMetPr 223
623 AGGAAA......AGGTCTCTGACAGATGCTGCAGGTGCT...TCACCTTC 663
223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
664 CAAGTACAGCCGTCAGTATGTCCGCATACCTGTGGAG..... 700
240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
701 ..AGTGGCCAAGCAATGGCTTTTTCTGAACATAATTCAGATGATGAGGAT 748
257 AlaValAlaSerSerGlyProIleIleSerAspIleThr.....Gl 270
749 GGA...AATCGTACAGGTCTCATCATCCGAGACATCACTGATACCCTGGA 795
270 uLeuAlaProAlaSerProMetAlaSerProGlyGlySerIleAspGluA 287
796 AAATGCCACCAATGGTCTCCTTGCTGTAGCACACACAAGTGGCAGAGACA 845
287 rgProLeuSerSerSer.....ProLeuValArgValLys 298
846 GAGAGACGCAGACAGCTCTGGATCCTGGCTTACCTATTTGTCAAGTATCA 895
299 Glu.....GluProProSerProProGlnSerProAr 309
896 CAGCCCAATGAGTTAAGCTGTGCAGAACCTATCCCTCCA......GTTCA 939
309 gValGluGluAlaSer...ProGlyArgProSerSerValAspThrLeuL 325
940 TATAAATGATGTCAGCAAGCCCAATGAAATGGGAAATGTTGCTGTAGAAC 989
325 eu.....SerProThrAlaLeuIleAsp 332
990 TGCACACTGCGCAAGCTAATGCTCCAGAAGACCCCGTGTCAGTGATTGAC 1039
333 SerIleLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuTh 349
1040 TCCATCTTGAATGAG..... 1054
349 rAspAlaArgGlyHisThrAspThrGluGlyArgProProSerProProP 366
1055 .....AATAACTCTGGAAACCAAAATGATCCT.... 1081
366 roThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsn 382
1082 .....TTACTGGACAGAGAA 1096
383 GluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGl 399
1097 GAAATTCAGGATTTTCTTAATTGCATTGATGCCAGCCTCGAAGAGCTTCA 1146
399 nThrMetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuA 416
1147 AGCAATGTTATCTGGGAAGCAGTATAGCTTTGGTTCTGAAGCTTTCAGCG 1196
416 spLeuPheSerProSerValThrValProAspMetSerLeuProAspLeu 432
1197 ATGTGTTTAACCCTGAACTGCCAGCCCTGGATATGAACTTGATGGAAACT 1246
433 AspSerSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProPr 449
1247 TCCCCTGGTATGGAAAATATTGCAAACATGGAAGACAGCACCGAAGATCT 1296
449 oArgProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuV 466
1297 GGGAGCAAGTGAGAGAGAAACAGCAGGAAGCAAAGGTGGGCAGGAA.... 1342
466 alHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAsp 482
1343 .....GGA 1345
483 ThrGlySerAsnAspLeuProValLeuPheGlu 493
1346 ACAGAGAGCTGTGACAGTTCCGTACTCTTCCAA 1378

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2000, 20:20:05 ; Search time 97.27 Seconds
(without alignments)
5545.534 Million cell updates/sec

Title: US-09-304-121-1
Perfect score: 2156
Sequence: 1 cgggcccgttgcaagatggc.....aaaaaaaaaaaaaaaaaaa 2156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2156	100.0	2156	1	V32958	Human wild-type he
2	2149.6	99.7	2156	1	Q13241	Human HSF cDNA seq
3	2133	98.9	2156	1	Q25713	Sequence of human
4	184.4	8.6	2781	1	Q13239	HSF cDNA sequence.
5	184.2	8.5	2781	1	Q25712	Sequence of Drosop
6	120.4	5.6	350	1	T84949	Human prostate pro
7	56.8	2.6	1362	1	V59101	S. aureofaciens te
8	51.4	2.4	114955	1	X53491	Human adenosine A1
c 9	51	2.4	114955	1	X53491	Human adenosine A1
10	48.8	2.3	3198	1	X02974	Human IL-1ra BAC c
11	48.6	2.3	1791	1	V82460	Triticum sp. cyste
12	48.6	2.3	2004	1	T85356	Nephila clavipes s
13	46.8	2.2	2338	1	Q14183	N.clavipes draglin
14	46.8	2.2	2338	1	V23249	Nephila clavipes s
15	46.4	2.2	2517	1	Q93037	Human E2F-1 comple
16	46.4	2.2	2517	1	V20440	Human EGF-1 oncoge
17	46.4	2.2	2517	1	V40003	Transcription fact
18	46.4	2.2	2517	1	X01354	E2F1 transcription
19	45.6	2.1	2888	1	T59268	Streptomyces prist
20	45.6	2.1	53789	1	V21187	Amycolatopsis medi
21	45.4	2.1	29879	1	Q46806	eryA region of S.
22	45.2	2.1	387	1	T58552	Streptomyces prist
23	44.4	2.1	2742	1	V49291	Aujeszky's disease
24	44.2	2.1	2403	1	T13955	Maize-optimized VI
25	44.2	2.1	2403	1	T74004	Maize optimised-B.
26	44.2	2.1	2403	1	V16177	Maize optimised DN
27	44.2	2.1	2403	1	V68065	Maize-optimised DN
28	44.2	2.1	3586	1	T46187	DNA encoding cell
29	44.2	2.1	3624	1	T46186	DNA encoding cell
30	44	2.0	1266	1	X09011	Brn-3a polynucleot
31	43.8	2.0	3600	1	T13227	Thermostable enzym
32	43.6	2.0	3474	1	Q39018	Sequence of pure m
33	43.6	2.0	3474	1	T74012	Maize optimised-B.
34	43.6	2.0	3474	1	V16191	Maize optimised DN
35	43.6	2.0	3508	1	Q39020	Sequence of full l
36	43.6	2.0	3508	1	T74013	Full length maize
37	43.6	2.0	3508	1	V16192	Maize optimised DN
38	43.4	2.0	2000	1	N71065	Gene encoding Plas
39	43.4	2.0	36519	1	V22141	Chimpanzee adenovi
40	43	2.0	1140	1	T88392	Corn barnacle G-pr
41	43	2.0	1766	1	V04490	Corn p-hydroxyphen
42	42.8	2.0	1913	1	Q15015	Nuclear factor C/E
43	42.8	2.0	2241	1	V68070	Maize-optimised DN
44	42.8	2.0	2796	1	N50036	Partial sequence o
45	42.8	2.0	30001	1	X05110	S. aureofaciens DN

ALIGNMENTS

RESULT 1
V32958
ID V32958 standard; DNA; 2156 BP.
AC V32958;
DT 26-OCT-1998 (first entry)
DE Human wild-type heat shock transcription factor 1 (HSF1) DNA.
KW Human wild-type heat shock transcription factor 1; HSF1; ischaemia;
KW heat shock protein; hsp; UV-B light; sepsis; hyperthermia;
KW oxidative stress; anti-tumour agent; cancer cell; cytotoxic agent; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 161. .1750
FT /*tag= a
FT /product= "HSF1"
PN WO9831803-A1.
PD 23-JUL-1998.
PF 21-JAN-1998; U01038.
PR 19-AUG-1997; US-914646.
PR 21-JAN-1997; US-035662.
PA (UYMI-) UNIV MIAMI.
PI Voellmy RW;
DR WPI; 98-414102/35.
DR P-PSDB; W49093.
PT Method for modulating synthesis of heat-shock protein - by
PT administering mutant heat shock transcription factors, used, e.g to
PT protect cells against chemotherapy
PS Disclosure; Fig 1A-1C; 84pp; English.
CC The present sequence represents the wild-type heat shock transcription
CC factor 1 (HSF1) DNA. The invention provides a method for modulating
CC the expression of endogenous heat shock protein (hsp) genes in
CC eukaryotic cells. The method involves introducing a mutated HSF1
CC or a gene encoding the mutated HSF1 into cells, with the result
CC that hsp synthesis in the recipient cells is altered. Mutation of
CC a regulatory region, spanning residues 180 to 315, in the HSF1
CC sequence resulted in hsp synthesis activation in the absence of
CC stress. This positive acting mutant HSF1 is claimed to induce a
CC protected state in the cell. Mutation of a second region, spanning
CC residues 277 to 529, in the HSF1 sequence resulted in hsp synthesis
CC inhibition induced by stress in the presence of stress. This negative
CC acting mutated HSF1 is claimed to induce a sensitised state in a cell.
CC The mutant HSF1s are claimed to be useful for protecting cells against
CC damage caused by therapeutics, UV-B light, sepsis, hyperthermia,
CC oxidative stress and ischaemia, particularly to increase resistance of
CC normal cells to anti-tumour agents, or to increase immunogenicity of
CC cancer cells. The mutant HSF1s are active in absence of stress,
CC unlike wild-type HSF1, even when over expressed, and eliminate the need
CC for cytotoxic agents for regulating the heat-shock system.
SQ Sequence 2156 BP; 435 A; 739 C; 628 G; 354 T;

Query Match 100.0%; Score 2156; DB 1; Length 2156;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 cgggcccgttgcaagatggcggcggccatgctgggccccggggctgtgtgtgcgcagcgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGGCCCGTTGCAAGATGGCGGCGGCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGG 60
```

```
Qy      61 gcggcggcgcggcccggaaggctggcgcggcgacggcgttagcccggccctcggcccctc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCGGCGGCGCGGCCCGGAAGGCTGGCGCGGCGACGGCGTTAGCCCGGCCCTCGGCCCCTC 120

Qy     121 tttgcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTTGCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCC 180

Qy     181 cggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgag 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGGCGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAG 240

Qy     241 cgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgtt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTT 300

Qy     301 cgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

Qy     361 cagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcaggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGG 420

Qy     421 cggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtgg 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGG 480

Qy     481 ccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagag 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAG 540

Qy     541 tgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgat 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGAT 600

Qy     601 gaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

Qy     661 tctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAA 720

Qy     721 caagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagag 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGAG 780

Qy     781 aaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCCG 840

Qy     841 gcagttctccctggagcacgtccacggctcgggccccctactcggccccctccccagccta 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTA 900

Qy     901 cagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctccga 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

Qy     961 catcaccgagctggctcctgccagccccatggcctccccggcgggagcatagacgagag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCATAGACGAGAG 1020

Qy    1021 gcccctatccagcagccccctggtgcgtgtcaaggaggagccccccagcccgcctcagag 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCCCCCAGCCCGCCTCAGAG 1080

Qy    1081 cccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtcccc 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCCATCTTCCGTGGACACCCTCTTGTCCCC 1140

Qy    1141 gaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GACCGCCCTCATTGACTCCATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGC 1200

Qy    1201 cctcacggacgccaggggccacacggacaccgagggccggcctccctccccccgcccac 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CCTCACGGACGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

Qy    1261 ctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgacca 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGCTCAGTGACCA 1320

Qy    1321 cttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacggctt 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACCATGCTGAGCAGCCACGGCTT 1380

Qy    1381 cagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacat 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACAT 1440

Qy    1441 gagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctccccagga 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GAGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGA 1500

Qy    1501 gccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgca 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GCCCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

Qy    1561 ctacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacga 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGA 1620

Qy    1621 cctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcgc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 CCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCGC 1680

Qy    1681 cgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccccac 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCAC 1740

Qy    1741 tgtctcctagaggccccggaggagctgggccagccgcccaccccccaccccccagtgcaggg 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGG 1800

Qy    1801 ctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgccata 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 CTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATA 1860

Qy    1861 gccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcaccct 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 GCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCCT 1920

Qy    1921 ggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttcacagc 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAGC 1980

Qy    1981 cacacctggactgaccctgcaggttgttcatagtcagaattgtattttggatttttacac 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 CACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACAC 2040

Qy    2041 aactgtcccgttccccgctccacagagatacacagatatatacacacagtggatggacgg 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 AACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACGG 2100

Qy    2101 acaagacaggcagagatctataaacagacaggctctaaaaaaaaaaaaaaaaaaaaa 2156
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 ACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAAA 2156
```

RESULT 2
Q13241
ID Q13241 standard; cDNA; 2156 BP.
AC Q13241;

```
DT   29-OCT-1991  (first entry)
DE   Human HSF cDNA sequence.
KW   Heat shock factor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             160. .1750
FT                   /*tag=  a
FT   polya_site      2137. .2156
FT                   /*tag=  b
PN   US7617901-A.
PD   16-JUL-1991.
PF   26-NOV-1990; 617901.
PR   26-NOV-1990; US-617901.
PA   (USSH ) NAT INST OF HEALTH.
PI   Wu C, Clos J, Westwood JT, Rabindran S;
DR   WPI; 91-252343/34.
DR   P-PSDB; R13503.
PT   DNA encoding Drosophila and human heat shock factor proteins -
PT   used for developing prods. for studying stress and disease states
PT   in living systems.
PS   Disclosure; Fig 13; 68pp; English.
CC   The sequence encodes human heat shock factor protein and was
CC   obtained by using short stretches of homologous sequences between
CC   Drosophila and yeast HSFs as primers for the polymerase chain
CC   reaction. The human HSF DNA was obtd. by screening human B cell
CC   cDNA libraries with the amplified sequence. The HSF sequence can be
CC   used to identify the HSF genes in other organisms and also for the
CC   detection of stress or a diseased state in living systems. The gene
CC   can be used to increase expression of other gene prods. by co-
CC   transfecting the HSF gene together with other genes linked to heat
CC   shock elements. It can be linked to a tissue-general or tissue
CC   specific promoter and introduced into transgenic mice as a tool for
CC   eliciting increased or chronic stress response conditions as a
CC   model for how tissues respond to chronic stress conditions such as
CC   those caused by viral infection, chemical or mechanical stress.
CC   There is no stop codon upstream of the designated AUG start codon
CC   so these sequences and the encoded amino acids, including thosee
CC   further upstream not yet sequenced may form part of the natural HSF
CC   protein. See also Q13240 and Q13241.
SQ   Sequence   2156 BP;    435 A;    743 C;    624 G;    354 T;
```

Query Match 99.7%; Score 2149.6; DB 1; Length 2156;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 cgggcccgttgcaagatggcggcggccatgctgggccccggggctgtgtgtgcgcagcgg 60
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      1 CGGGCCCGTTGCAAGATGGCGGCGCCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGG 60

Qy     61 gcggcggcgcggcccggaaggctggcgcggcgacggcgttagcccggccctcggcccctc 120
          |||||| ||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db     61 GCGGCCGCGCGGCCCGGAAGGCTGGCGCCCCGACGGCGTTAGCCCGGCCCTCGGCCCCTC 120

Qy    121 tttgcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCC 180

Qy    181 cggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGGCGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAG 240

Qy    241 cgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgtt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTT 300

Qy    301 cgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

Qy    361 cagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcaggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGG 420

Qy    421 cggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGG 480

Qy    481 ccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagag 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAG 540

Qy    541 tgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGAT 600

Qy    601 gaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

Qy    661 tctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAA 720

Qy    721 caagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGAG 780

Qy    781 aaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCCG 840

Qy    841 gcagttctccctggagcacgtccacggctcgggcccctactcggcccccctccccagccta 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTA 900

Qy    901 cagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctccga 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

Qy    961 catcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacgagag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCATAGACGAGAG 1020

Qy   1021 gcccctatccagcagccccctggtgcgtgtcaaggaggagccccccagcccgcctcagag 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCCCCCAGCCCGCCTCAGAG 1080

Qy   1081 cccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtcccc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCCATCTTCCGTGGACACCCTCTTGTCCCC 1140

Qy   1141 gaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GACCGCCCTCATTGACTCCATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGC 1200

Qy   1201 cctcacggacgccaggggccacacggacaccgagggccggcctccctccccccgcccac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTCACGGACGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

Qy   1261 ctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgacca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGCTCAGTGACCA 1320

Qy   1321 cttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacggctt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACCATGCTGAGCAGCCACGGCTT 1380

Qy   1381 cagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACAT 1440

Qy   1441 gagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctccccagga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GAGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGA 1500
```

```
Qy   1501 gccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

Qy   1561 ctacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacga 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGA 1620

Qy   1621 cctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcgc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCGC 1680

Qy   1681 cgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccccac 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCAC 1740

Qy   1741 tgtctcctagaggccccggaggagctgggccagccgcccaccccccaccccagtgcaggg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGG 1800

Qy   1801 ctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgccata 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATA 1860

Qy   1861 gccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcaccct 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCCT 1920

Qy   1921 ggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttcacagc 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAGC 1980

Qy   1981 cacacctggactgaccctgcaggttgttcatagtcagaattgtatttttggatttttacac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACAC 2040

Qy   2041 aactgtcccgttccccgctccacagagatacacagatatatacacacagtggatggacgg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACGG 2100

Qy   2101 acaagacaggcagagatctataaacagacaggctctaaaaaaaaaaaaaaaaaaaaa 2156
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAAA 2156
```

```
RESULT   3
Q25713
ID   Q25713 standard; cDNA; 2156 BP.
AC   Q25713;
DT   28-DEC-1992  (first entry)
DE   Sequence of human Heat Shock Factor (HSF) cDNA
KW   Heat shock factor; stress condition; assay; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             160. .1750
FT                   /*tag= a
PN   WO9209617-A.
PD   11-JUN-1992.
PF   22-NOV-1991; U08592.
PR   26-NOV-1990; US-617910.
PA   (USDC ) US DEPT OF COMMERCE.
PI   Clos J, Rabindran S, Westwood JT, Wu C;
DR   WPI; 92-217013/26.
DR   P-PSDB; R24948.
PT   DNA fragment encoding Drosophila or human heat shock factor
PT   protein - and use of corresp. monoclonal antibodies for
PT   diagnosing abnormal stress conditions in cells
PS   Claim 5; Figure 13; 75pp; English.
CC   The cloning of human heat shock factor HuHSF) was achieved by using
CC   short stretches of homologous sequences between Drosophila and
CC   yeast heat shock factors as primers in the polymerase chain
CC   reaction (PCR) (Q25714,Q25715). The HuHSF length clone was obtained
CC   by screening human cDNA libraries with the amplified sequence. The
CC   HuHSF cDNA clone includes an open reading frame of 529 AAs with a
CC   calculated molecular weight of 58,000 (Q24713,R24948). The size of
CC   HuHSF as measured by SDS-polyacrylamide gel electrophoresis is
CC   60,000 which is in close agreement with the calculated size. The
CC   claims refer to Figure 12, rather than Figure 13, but this would
CC   appear to be an error in the claims.
SQ   Sequence   2156 BP;    435 A;    739 C;    628 G;    354 T;
```

```
  Query Match            98.9%;  Score 2133;  DB 1;  Length 2156;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 2155;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy      1 cgggcccgttgcaagatggcggcggccatgctgggccccggggctgtgtgtgcgcagcgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGGCCCGTTGCAAGATGGCGGCGGCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGG 60

Qy     61 gcggcggcgcggcccggaaggctggcgcggcgacggcgttagcccggccctcggcccctc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGGCGGCGCGGCCCGGAAGGCTGGCGCGGCGACGGCGTTAGCCCGGCCCTCGGCCCCTC 120

Qy    121 tttgcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCC 180

Qy    181 cggcgcggcgggg-cccagcaacgtcccggccttcctgaccaagctgtggaccctcgtga 239
          |||||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db    181 CGGCGCGGCGGGGCCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGA-CCTCGTGA 239

Qy    240 gcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgt 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 GCGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGT 299

Qy    300 tcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatgg 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 TCGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGG 359

Qy    360 ccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcagg 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 CCAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGG 419

Qy    420 gcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtg 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 GCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTG 479

Qy    480 gccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaaga 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 GCCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGA 539

Qy    540 gtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctga 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GTGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGA 599

Qy    600 tgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgagg 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 TGAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGG 659

Qy    660 ctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtca 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 CTCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCA 719

Qy    720 acaagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaaga 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 ACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGA 779

Qy    780 gaaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagcc 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 GAAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCC 839

Qy    840 ggcagttctccctggagcacgtccacggctcgggcccctactcggcccccctccccagcct 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 GGCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCT 899
```

```
Qy    900 acagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctccg 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 ACAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCG 959

Qy    960 acatcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacgaga 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 ACATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCATAGACGAGA 1019

Qy   1020 ggcccctatccagcagccccctggtgcgtgtcaaggaggagcccccccagcccgcctcaga 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 GGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCCCCCAGCCCGCCTCAGA 1079

Qy   1080 gccccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtccc 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 GCCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCCATCTTCCGTGGACACCCTCTTGTCCC 1139

Qy   1140 cgaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacag 1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 CGACCGCCCTCATTGACTCCATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAG 1199

Qy   1200 ccctcacggacgccaggggccacacggacaccgagggccggcctccctcccccccgccca 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 CCCTCACGGACGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCA 1259

Qy   1260 cctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgacc 1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 CCTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGCTCAGTGACC 1319

Qy   1320 acttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacggct 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 ACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACCATGCTGAGCAGCCACGGCT 1379

Qy   1380 tcagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgaca 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 TCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACA 1439

Qy   1440 tgagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctccccagg 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 TGAGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGG 1499

Qy   1500 agccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgc 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500 AGCCCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGC 1559

Qy   1560 actacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacg 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 ACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACG 1619

Qy   1620 acctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcg 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 ACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCG 1679

Qy   1680 ccgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggacccca 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 CCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCA 1739

Qy   1740 ctgtctcctagaggccccggaggagctgggccagccgcccaccccaccccagtgcagg 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 CTGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGG 1799

Qy   1800 gctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgccat 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 GCTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCAT 1859

Qy   1860 agccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcaccc 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 AGCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCC 1919

Qy   1920 tggcctgccagtctgccttccccaaccccgtgtcctgtggtttggttggggcttcacag 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 TGGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAG 1979

Qy   1980 ccacacctggactgaccctgcaggttgttcatagtcagaattgtattttggatttttaca 2039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 CCACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACA 2039

Qy   2040 caactgtcccgttccccgctccacagagatacacagatatatacacacagtggatggacg 2099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040 CAACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACG 2099

Qy   2100 gacaagacaggcagagatctataaacagacaggctctaaaaaaaaaaaaaaaaaaaaa 2156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100 GACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAAA 2156
```

RESULT 4
Q13239
ID Q13239 standard; cDNA; 2781 BP.
AC Q13239;
DT 29-OCT-1991 (first entry).
DE HSF cDNA sequence.
KW Heat shock factor; ss.
OS Drosophila.

FH	Key	Location/Qualifiers
FT	cds	229. .2304
FT		/*tag= a
FT	polya_signal	2722. .2727
FT		/*tag= b
FT	polya_site	1757. .1781
FT		/*tag= c

PN US7617901-A.
PD 16-JUL-1991.
PF 26-NOV-1990; 617901.
PR 26-NOV-1990; US-617901.
PA (USSH) NAT INST OF HEALTH.
PI Wu C, Clos J, Westwood JT, Rabindran S;
DR WPI; 91-252343/34.
DR P-PSDB; R13502.
PT DNA encoding Drosophila and human heat shock factor proteins -
PT used for developing prods. for studying stress and disease states
PT in living systems.
PS Disclosure; Fig 2; 68pp; English.
CC The sequence encodes Drosophila heat shock factor protein and was
CC obtained by screening a Drosophila genomic library with oligo-
CC nucleotide probes (Q13237, Q13238) based on the HSF amino acid
CC sequence. The HSF sequence can be used to identify the HSF genes in
CC other organisms and also for the detection of stress or a diseased
CC state in living systems. The gene can be used to increase
CC expression of other gene prods. by cotransfecting the HSF gene
CC together with other genes linked to heat shock elements. It can be
CC linked to a tissue-general or tissue-specific promoter and
CC introduced into transgenic mice as a tool for eliciting increased
CC or chronic stress response conditions as a model for how tissues
CC respond to chronic stress conditions such as those caused by viral
CC infection, chemical or mechanical stress. See also Q13240 and
CC Q13241.
SQ Sequence 2781 BP; 831 A; 631 C; 690 G; 629 T;

Query Match 8.6%; Score 184.4; DB 1; Length 2781;
Best Local Similarity 59.0%; Pred. No. 1.1e-27;
Matches 374; Conservative 0; Mismatches 251; Indels 9; Gaps 3;

```
Qy    176 ggccccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctc 235
          ||   || ||||||       |||   || ||||||||  || ||||  |||||  |||
Db    337 GGAGACGCCGCGGCCATCGGAAGCGGGGTGCCGGCCTTTTTGGCCAAATTGTGGCGCCTG 396

Qy    236 gtgagcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccac 295
          |||  |||  | || ||| |     | || |||||||  |   |   ||  | || |||
Db    397 GTGGACGATGCCGATACCAATCGCTTGATTTGCTGGACCAAGGATGGCCAAAGTTTCGTT 456

Qy    296 gtgttcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaac 355
          |     | || |  || ||||||||||  |  ||||     ||| ||||||||||||
Db    457 ATTCAAAATCAAGCGCAATTTGCCAAGGAACTATTGCCACTAAACTACAAGCACAACAAC 516

Qy    356 atggccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgag 415
```

```
Db    517 ATGGCCAGTTTCATAAGGCAATTGAATATGTATGGATTCCACAAGATCACCTCTATTGAC 576

Qy    416 cagggcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctg 475
Db    577 AATGGCGGAC---TACGTTTTGATCGCGACGAGATTGAATTTTCGCACCCATTTTTTAAG 633

Qy    476 cgtggccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctg 535
Db    634 CGCAACTCGCCTTTTCTACTTGACCAAATCAAAAGGAAAATATCGAACAACAAAAATGGT 693

Qy    536 aagagtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcag 595
Db    694 GACGACAAAGGTGTCCTGAAGCCGGAGGCCA---TGTCGAAGATTCTCACCGATGTGAAA 750

Qy    596 ctgatgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaat 655
Db    751 GTCATGCGGGGTCGTCAGGACAATCTGGATTCGCGCTTCTCCGCCATGAAACAGGAGAAC 810

Qy    656 gaggctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtc 715
Db    811 GAAGTGCTGTGGCGCGAGATAGCCAGCCTGCGCCAAAAGCACGCTAAGCAGCAACAAATA 870

Qy    716 gtcaacaagctcattcagttcctgatctcactggtgcagtcaaaccggatcctg---ggg 772
Db    871 GTCAACAAACTGATCCAGTTCCTCATTACCATTGTGCAACCGTCGCGCAACATGTCTGGC 930

Qy    773 gtgaagagaaagatccccctgatgctgaacgaca 806
Db    931 GTGAAACGCCATGTGCAGCTGATGATGAACAATA 964
```

RESULT 5
Q25712
ID Q25712 standard; cDNA; 2781 BP.
AC Q25712;
DT 28-DEC-1992 (first entry)
DE Sequence of Drosophila heat shock factor (HSF) cDNA.
KW Heat shock factor; stress condition; assay; ss.
OS Drosophila.
FH Key Location/Qualifiers
FT cds 229. .2315
FT /*tag= a
FT polya_signal 2723. .2728
FT /*tag= b
PN WO9209617-A.
PD 11-JUN-1992.
PF 22-NOV-1991; U08592.
PR 26-NOV-1990; US-617910.
PA (USDC) US DEPT OF COMMERCE.
PI Clos J, Rabindran S, Westwood JT, Wu C;
DR WPI; 92-217013/26.
DR P-PSDB; R24947.
PT DNA fragment encoding Drosophila or human heat shock factor
PT protein - and use of corresp. monoclonal antibodies for
PT diagnosing abnormal stress conditions in cells
PS Claim 3; Fig 2B; 75pp; English.
CC Two 20-mer oligonucleotides with 32-fold degeneracy (Q25710,Q25711),
CC based on the predicted nucleotide sequences of HSF peptide 27 and
CC peptide 29 were used to probe a Drosophila genomic library.
CC Initially two genomic DNA clones were identified which contained a
CC common, 1800nt SalI-EcoRI fragment. This fragment, which hybridised
CC with both oligo probes, was then used to isolate cDNA clones from a
CC random-primed and an oligo dT-primed cDNA library. The 2.8 kb of HSF
CC cDNA sequence reconstructed from six overlapping cDNA clones reveals
CC a single open reading frame of 691 amino acids. The sequences of all
CC six HSF tryptic peptides within the 691-amino acid open reading frame
CC were located, and thus concluded that this reading frame encodes
CC Drosophila HSF (Q25712,R24947). The molecular mass of Drosophila
CC HSF, calculated from the deduced amino acid sequence is 77,300
CC daltons, significantly lower than the apparent mass of 110,000
CC daltons measured by SDS gel electrophoresis. Evidently, Drosophila
CC, HSF has an anomalous mobility on SDS gels. Fig 2B (Q25712) has a
CC non-standard nt (D) at posn. 741.
SQ Sequence 2781 BP; 830 A; 639 C; 682 G; 629 T;

Query Match 8.5%; Score 184.2; DB 1; Length 2781;
Best Local Similarity 57.8%; Pred. No. 1.2e-27;
Matches 366; Conservative 1; Mismatches 260; Indels 6; Gaps 2;

```
Qy    176 ggccccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctc 235
Db    337 GGAGACGCCGCGGCCATCGGAAGCGGGGTGCCGGCCTTTTGGCCAAATTGTGGCGCCTG 396

Qy    236 gtgagcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccac 295
Db    397 GTGGACGATGCCGATACCAATCGCTTGATTTGCTGGACCAAGGATGGCCAAAGTTTCGTT 456

Qy    296 gtgttcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaac 355
Db    457 ATTCAAAATCAAGCGCAATTTGCCAAGGAACTATTGCCACTAAACTACAAGCACAACAAC 516

Qy    356 atggccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgag 415
Db    517 ATGGCCAGTTTCATAAGGCAATTGAATATGTATGGATTCCACAAGATCACCTCTATTGAC 576

Qy    416 cagggcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctg 475
Db    577 AATGGCGGAC---TACGTTTTGATCGCGACGAGATTGAATTTTCGCACCCATTTTTTAAG 633

Qy    476 cgtggccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctg 535
Db    634 CGCAACTCGCCTTTTCTACTTGACCAAATCAAAAGGAAAATATCGAACAACAAAAATGGT 693

Qy    536 aagagtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcag 595
Db    694 GACGACAAAGGTGTCCTGAAGCCGGAGGCCA---TGTCGAAGATTCTCACDGATGTGAAA 750

Qy    596 ctgatgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaat 655
Db    751 GTCATGCGGGGTCGTCAGGACAATCTGGATTCGCGCTTCTCCGCCATGAAACAGGAGAAC 810

Qy    656 gaggctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtc 715
Db    811 GAAGTGCTGTGGCGCGAGATAGCCAGCCTGCGCCAAAAGCACGCTAAGCAGCAACAAATA 870

Qy    716 gtcaacaagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtg 775
Db    871 GTCAACAAACTGATCCAGTTCCTCATTACCATTGTGCAACCGTCGCGCAACATGTCTGGC 930

Qy    776 aagagaaagatccccctgatgctgaacgacagt 808
Db    931 GTGAAACGCCATGTGCAGCTGATGATCAACAAT 963
```

RESULT 6
T84949
ID T84949 standard; cDNA; 350 BP.
AC T84949;
DT 27-APR-1998 (first entry)
DE Human prostate protein HPA38 cDNA.
KW Prostate cancer; immunotherapy; therapy; immunodiagnosis; diagnosis;
KW vaccine; human; HPA38; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 3. .350
FT /*tag= a
PN WO9733909-A2.
PD 18-SEP-1997.
PF 14-MAR-1997; U04192.
PR 11-APR-1996; US-633840.
PR 15-MAR-1996; US-616745.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Reed SG, Twardzik DR;
DR WPI; 97-470816/43.

```
DR   P-PSDB; W27306.
PT   Immunogenic portions of prostate proteins - useful to develop
PT   products to detect, monitor, treat or inhibit development of
PT   prostate cancer
PS   Claim 28; Page 69; 84pp; English.
CC   This cDNA sequence includes a coding region for human prostate
CC   protein HPA38 (see W27306), an immunogenic portion of which can be
CC   used in a claimed pharmaceutical composition for the treatment of
CC   prostate cancer, in a claimed vaccine for treatment of prostate
CC   cancer, or to raise claimed antibodies suitable for use in
CC   diagnosis or monitoring the progression of prostate cancer.  HPA38
CC   cDNA was isolated from a human prostate adenocarcinoma cell line
CC   LnCap.fgc (ATCC CRL-1740) cell cDNA library by expression screening
CC   with human prostatitis sera.  DNA sequences (see T84927-52) for 17
CC   HPA proteins (see W23312-23 and W27303-07) are claimed and can be
CC   used to produce recombinant HPA polypeptides in host cells
CC   (particularly E. coli, yeast and mammalian cell lines) and to
CC   design primers and probes for use in claimed methods of detecting
CC   prostate cancer.
SQ   Sequence   350 BP;    102 A;    77 C;    82 G;    89 T;


  Query Match             5.6%;  Score 120.4;  DB 1;  Length 350;
  Best Local Similarity   65.0%;  Pred. No. 2.5e-15;
  Matches  178;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy     200 aacgtcccggccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcg 259
           ||||| ||||| ||||| | |||||||||||| || |||   ||  |  ||||  ||| |
Db      75 AACGTGCCGGCTTTCCTCAGCAAGCTGTGGACGCTTGTGGAGGAAACCCACACTAACGAG 134

Qy     260 ctcatctgctggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgcc 319
           |||||  |||||||| ||  ||  | || || |  || || ||  ||   |  |||||
Db     135 TTCATCACCTGGAGCCAGAATGGCCAAAGTTTTCTGGTCTTGGATGAGCAACGATTTGCA 194

Qy     320 aaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctc 379
           || ||  | || ||||| || |||||||||| || ||||| ||||| ||| |||| ||
Db     195 AAAGAAATTCTTCCCAAATATTTCAAGCACAATAATATGGCAAGCTTTGTGAGGCAACTG 254

Qy     380 aacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagag 439
           || |||||||| ||||| ||||| | || |||||    ||     |    |   |
Db     255 AATATGTATGGTTTCCGTAAAGTAATACATATCGACTCTGGAATTGTTAAGCAAGAAAGA 314

Qy     440 agagacgacacggagttccagcacccatgcttcc 473
               |       || || ||||| || | |||||
Db     315 GATGGTCCTGTAGAATTTCAGCATCCTTACTTCC 348


RESULT   7
V59101
ID   V59101 standard; DNA; 1362 BP.
AC   V59101;
DT   20-JAN-1999  (first entry)
DE   S. aureofaciens tetracycline dehydrogenase gene.
KW   Tetracycline dehydrogenase; tetracycline production; antibiotic;
KW   6-Demethyltetracycline; DMT; ds.
OS   Streptomyces aureofaciens.
FH   Key             Location/Qualifiers
FT   CDS             190. .1359
FT                   /*tag=  a
PN   J10286091-A.
PD   27-OCT-1998.
PF   15-APR-1997; 097232.
PR   15-APR-1997; JP-097232.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
DR   WPI; 99-017003/02.
DR   P-PSDB; W73165.
PT   New DNA participating in biosynthesis of tetracycline - used to
PT   transfect the transgenic host and produce the antibiotic
PS   Claim 1; Page 9-11; 14pp; Japanese.
CC   This sequence represents the DNA of the invention and encodes the
CC   Streptomyces aureofaciens tetracycline dehydrogenase. A recombinant host
CC   cell containing this sequence can be used to produce the tetracycline
CC   antibiotic. 6-Demethyltetracycline (DMT) productivity can be improved
CC   using the DNA fragment.
SQ   Sequence   1362 BP;    159 A;    511 C;    528 G;    164 T;


  Query Match             2.6%;  Score 56.8;  DB 1;  Length 1362;
  Best Local Similarity   46.1%;  Pred. No. 0.0078;
  Matches  227;  Conservative    0;  Mismatches  262;  Indels    3;  Gaps    1;

Qy     892 cccagcctacagcagctccagcctctacgcccctgatgctgtggccagctctggacccat 951
           | | ||| || ||  | ||    || |||||      |    ||    | |  | | |
Db     147 CGCCGCCGACCGCGACGCCGCGATCGACGCCGAGACCGTCCGGGTGGTCGCCCGGCGCGG 206

Qy     952 catctccgacatcaccgagctggctcctgccagccccatggcctcccccggcgggagcat 1011
               ||   |||  ||||         | |  |  ||||||| || |||| |    |
Db     207 ACCGGCCCGGATCGTCGAGAACCGCAACGGCTTCGTCATGGCCGCCGCCGGGGTCGACGC 266

Qy    1012 agacgagaggcccctatccagcagccccctggtgcgtgtcaaggaggagcccccagccc 1071
           | | |   |||    || | |  ||| |||  |   |   |||  || | |   |
Db     267 CTCCAACACCGCCCCCGGCACCGTGCTGCTGCTGCCCGAGGACCCGGACGCCTCGGCGCG 326

Qy    1072 gcctcagagcccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccct 1131
           | | | | || ||||| |  || ||  ||   ||||| |||       |||| | |  |
Db     327 GGCGCTGCGCTCCCGGCTCCAGCAGCTGA---CCGGGTGCCGGCTCGCCGTCGTCGTCAC 383

Qy    1132 cttgtccccgaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctc 1191
           |    ||          | |      |  |   ||    ||    |     | | |||||
Db     384 CGACACCTTCGGACGGCCCTGGCGCAACGGGCTCACCGACGTCGCGATCGGCGCCGCCGG 443

Qy    1192 cgtcacagccctcacggacgccaggggccacacggacaccgagggccggcctccctcccc 1251
             |  | | ||||  ||||  | |||| | ||| |||| | | |||      |     |
Db     444 GCTGTCCGTCCTCGAGGACCACCGGGGGCGCACCGACAGCCACGGCAACGAGCTGGTGCT 503

Qy    1252 cccgcccacctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagct 1311
           |   |||| | ||| | ||  |   |  |  ||  |||  |||||  |||    ||
Db     504 GACCGTCACCGCGACCGCCGACGAACTCGCCGCCGCCGCCGACCTGGTGAAGGGCAAGGC 563

Qy    1312 cagtgaccacttggatgctatggactccaacctggataacctgcagaccatgctgagcag 1371
           ||  | |  |  ||  || || |    |  ||| |   |||||   |||     |  | |
Db     564 CACCGGCGTCCCGGTCGCGATCGTGCGCGGCCTCGGACACCTGGTCACCGCCGAGGACGG 623

Qy    1372 ccacggcttcag 1383
           |  ||||  | |
Db     624 CGCCGGCACCCG 635


RESULT   8
X53491
ID   X53491 standard; DNA; 114955 BP.
AC   X53491;
DT   05-JUL-1999  (first entry)
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
OS   Synthetic.
PN   WO9913886-A1.
PD   25-MAR-1999.
PF   17-SEP-1998; U19419.
PR   09-JUN-1998; US-093972.
PR   17-SEP-1997; US-059160.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Nyce JW;
DR   WPI; 99-229400/19.
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
PS   Disclosure; Page 37; 120pp; English.
```

```
CC   The specification describes antisense oligonucleotides (X52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences X55272-74. These multiple target
CC   oligonucleotides (specifically X55180-271) can be used for the antisense
CC   treatment of diseases and conditions. Typical diseases and conditions
CC   are those associated with impaired respiration and inflammation,
CC   including lung diseases, pulmonary vasoconstriction, inflammation,
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC   hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC   metastases, as well as all types of cancers which may metastasize or have
CC   metastasized to the lungs, including breast and prostate cancer.
SQ   Sequence   114955 BP;  6071 A;   29417 C;   36712 G;   21328 T;

  Query Match             2.4%;  Score 51.4;  DB 1;  Length 114955;
  Best Local Similarity   30.1%;  Pred. No. 0.17;
  Matches  350;  Conservative  104;  Mismatches  697;  Indels   11;  Gaps    3;

Qy    838 ccggcagttctccctggagcacgtccacggctcgggcccctactcggccccctccccagc 897
          | |||       ||  : || :|  | || | | ||||       :  || |:  || |
Db 104336 CGGGCSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCCGGGSNNNDNNCCGCBGGCCBGG 104395

Qy    898 ctacagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctc 957
             |  |  | | :   :  || |:|  || : |||  ||    | ||     :  |
Db 104396 GCGCGCCGCCGGCCGGSNNNDNNCCGCBG--GCCBGGGCGCGCCGCCGGCCGSNNNDNNC 104453

Qy    958 cgacatcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacga 1017
          || :  | : |    | | || |||         ||:  ||: |||| |      | :
Db 104454 CGCBGGCCBGGGCGCGCCGCCGGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCSNN 104513

Qy   1018 gaggcccctatccagcagccccctggtgcgtgtcaaggaggagccccccagcccgcctca 1077
           :  || |:  || |    | ||    |:   :    | :|  |:   |   ||||| :
Db 104514 NDNNCCGCBGGCCBGGGCGCGCCGCCGGSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGSN 104573

Qy   1078 gagcccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtc 1137
            :  ||  :|    ||  |||   ||:      ||  :  ||: ||    ||
Db 104574 NNDNNCCGCBGGCCBGGGCGCGCCGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCSNNND 104633

Qy   1138 cccgaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcac 1197
            |  |:| ||:    |   |     :   | : |       || ||:   :      :
Db 104634 NNCCGCBGGCCBGGGCGCGCCGSNNNDNNCCGCBGGCCBGGGCGCGCCSNNNDNNCCGCB 104693

Qy   1198 agccctcacggacgccaggggccacacggacaccgagggccggcctccctccccccccgcc 1257
           |||:   ||  |:   :  |  :       || |:   :  |  |:  ||:     |
Db 104694 GGCCBGGGCGCGCSNNNDNNCCGCBGGCCBGGGCGCGSNNNDNNCCGCBGGCCBGGGCGC 104753

Qy   1258 cacctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtga 1317
          :   :   | |: |    |   :   :    ||: |||:|| |    :      : |
Db 104754 SNNNDNNCCGCBGGCCBGGGCGSNNNDNNCCGCBGGCCBGGGCSNNNDNNCCGCBGGCC- 104813

Qy   1318 ccacttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacgg 1377
                :|  :   :   | |:  ||:||       |       :        || |:|
Db 104813 ------BGGGSNNNDNNCCGCBGGCCBGGSNNNDNNCCGCBGGCCBGSNNNDNNCCGCBG 104866

Qy   1378 cttcagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccga 1437
             :    :   | |: |  :   :   | :    :  || |  ||  | ||| ||| :
Db 104867 GCCBSNNNDNNCCGCBGGCCSNNNDNNCGCBGGCCBGGGCGCGCCGCCGGCCGGGCCGSN 104926

Qy   1438 catgagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctcccca 1497
             :  ||: |||:|  |  | |  | ||| ||||  :       : || :  |    ||
Db 104927 NNDNNGCBGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNCBGGCCBGGGCGCGCCGC 104986

Qy   1498 ggagccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggt 1557
          .  .  . |  |   || :   :  :|   : |           | |   | :   :   || :|
```

```
Db 104987 CGGCCGGGCCGSNNNDNNBGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNGGCCBGG 105046

Qy   1558 gcactacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaa 1617
          || |  | | |  | | |          || :|  | || | |||      |||:
Db 105047 GCGCGCCGCCGGCCGGGCCGSNNNDNNGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNN 105106

Qy   1618 cgacctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggctt 1677
          |       ||  || |   | || |:   :   :     |  | |||   ||| ||:
Db 105107 CCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNCBGGGCGCGCCGCCGGCCGGGCCGSNNN 105166

Qy   1678 cgccgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccc 1737
             :| |  | || ||    ||  || | |     :   || ||  | |    |      :
Db 105167 DNNBGGGCGCGCCGCC--GGCCGGGCCGSNNNDNNGGGCGCGCCGCCGGCCGGGCCGSNN 105224

Qy   1738 cactgtctcctagaggccccggaggagctgggccagccgcccaccccccaccccccagtgca 1797
           :  | | |   |  | || ||  |    :   |   |  ||  ||   || :   :
Db 105225 NDNNGGCGCGCCGCCGGCCGGGCCGSNNNDNNGCGCGCCGCCGGCCGGGCCGSNNNDNNC 105284

Qy   1798 gggctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgcc 1857
          | || |      |||  |:   :  ||  ||| | |  || | :   :    | ||||
Db 105285 GCGCCGCCGGCCGGGCCGSNNNDNNGCGCCGCCGGCCGGGCCGSNNNDNNCGCCGCCGGC 105344

Qy   1858 atagccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcac 1917
             ||| :   :   |   ||| | |   :   :   | ||    |    |    :  |
Db 105345 CGGGCCGSNNNDNNGCCGCCGGCCGGGCCGSNNNDNNCCGCCGGCCGGGCCGSNNNDNNC 105404

Qy   1918 cctggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttcac 1977
           | |||| | | |        ||  |    |||        |   ||   |:   :  |
Db 105405 GCCGGCCGGGCCGSNNNDNNGCCGGCCGGGCCGSNNNDNNCCGGCCGGGCCGSNNNDNNC 105464

Qy   1978 agccacacctggactgaccctg 1999
          |||   || :   :    | |
Db 105465 GGCCGGGCCGSNNNDNNGGCCG 105486


RESULT   9
X53491/c
ID   X53491 standard; DNA; 114955 BP.
AC   X53491;
DT   05-JUL-1999  (first entry)
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
OS   Synthetic.
PN   WO9913886-A1.
PD   25-MAR-1999.
PF   17-SEP-1998; U19419.
PR   09-JUN-1998; US-093972.
PR   17-SEP-1997; US-059160.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Nyce JW;
DR   WPI; 99-229400/19.
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
PS   Disclosure; Page 37; 120pp; English.
CC   The specification describes antisense oligonucleotides (X52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences X55272-74. These multiple target
CC   oligonucleotides (specifically X55180-271) can be used for the antisense
CC   treatment of diseases and conditions. Typical diseases and conditions
```

```
CC   are those associated with impaired respiration and inflammation,
CC   including lung diseases, pulmonary vasoconstriction, inflammation,
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC   hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC   metastases, as well as all types of cancers which may metastasize or have
CC   metastasized to the lungs, including breast and prostate cancer.
SQ   Sequence   114955 BP;    6071 A;    29417 C;    36712 G;    21328 T;


  Query Match             2.4%;  Score 51;  DB 1;  Length 114955;
  Best Local Similarity   33.1%;  Pred. No. 0.21;
  Matches  351;  Conservative   69;  Mismatches  623;  Indels   18;  Gaps    4;

Qy     885 ccccctccccagcctacagcagctccagcctctacgcccctgatgctgtggccagctctg 944
           |||| ||| | |  | | |    | | | |    ||||||  |        | | |  |||
Db 111340 CCCCTTCCTCTGGGTGCTGTTCGTGCTGGCCACCCGCCCCCAAGTGGAGTGGCTGGGCTG 111281

Qy     945 gacccatcatctccgacatcaccgagctggctcctgccagccccatggcctcccccggcg 1004
           |      || : :: ||   |||| || | | :  :| |: :|     :   : ||| ||
Db 111280 GCGGGCCCACVCVVAACGGGACCGCGCCGACAVGCVCAAVGVCACCNNHNNNSGCCGACG 111221

Qy    1005 ggagcatagacgagaggcccctatccag--cagcccccctggtgcgtgtcaaggaggagcc 1062
             ||    | || | | || |  :   :  | |  || || |||          :|
Db 111220 CCAGAGCCGCCGCGTGCCCGCNNHNNNSCCCGACGGCCCGGAGCGGACGCCCNNHNNNSC 111161

Qy    1063 ccccagcccgcctcagagcccccgggtagaggaggcgagtcccgggcgcccatcttccgt 1122
            | | ||| |     |||   |      :     ||  | ||::  :|||   |  | |
Db 111160 GCGCGGCCGGGGCGGGAGAGACNNHNNNSGCCCAGCCCGCCCVVCCVGCCGGCCGCCGGG 111101

Qy    1123 ggacaccctcttgtcc---ccgaccgccctcattgactccatcctgcgggagagtgaacc 1179
             :   :| |  | ||   ||| ||:||  |  :    ||| |  |:||   ||   :
Db 111100 NNHNNNSCGCGGGACCVGGCCGGCCVCCGGCNNHNNNSCCAGCGCGVGGAGCAGCNNHNN 111041

Qy    1180 tgcccccgcctccgtcacagccctcacggacgccaggggccacacggacaccgagggccg 1239
            : || || |   | ||     | |      :|| |  ||  | ||  | | :  :||
Db 111040 NSGCCGCGACAAGGCCAAGCGGCGCNNHNNNSCCGGCAGCGGCGCGAGCGCNNHNNNSCG 110981

Qy    1240 gcctc-cctcccccccgcccacctccacccctgaaaagtgcctcagcgtagcctgcctgg 1298
            | |        ||||  |  | || |||  :   :  || |  |    || ||| ||
Db 110980 CCGGCAAGAGGGGCCCGGACCGCGGCAGCCCNNHNNNSCCCCGCGCCCGGCCCCGCCCGG 110921

Qy    1299 acaagaatgagctcagtgaccacttggatgctatggactccaacctggataacctgcaga 1358
                :    | | |   || |   |  ||       :        ||       ||||
Db 110920 GAGNNHNNNSTCGCCGGAGCCGCCGCGCAGCAGCCCNNHNNNSGGCGGCCGCAGGGCAGC 110861

Qy    1359 ccatgctgagcagccacggcttcagcgtggacaccagtgccctgctggacctgttcagcc 1418
            |||  | |  |  :   :||    ||  |:   || | || |  :     |  |  | ||:
Db 110860 CCACCCCGCCCNNHNNNSGCCCGCGCAGGVGGGCCCGVGCGCNNHNNNSCGCGCCCGGCV 110801

Qy    1419 cctcggtgaccgtgcccgacatgagcctgcctgaccttgacagcagcctggccagtatcc 1478
            |  :|: | ||| ||       | ||  ||| |||:  | |  | ||   ||     :
Db 110800 CGCVGVVGCCCGCGCNNHNNNSGCGCGAGCCCGACVGGGGCGACGGCGACGCGCCNNHNN 110741

Qy    1479 aagagctcctgtctccccaggagccccccaggcctcccgagg-----------cagaga 1526
              | ||  |  | | : ||  :  || ||       ||||  |           |
Db 110740 NSGGGCCCGGGCCCVGCCGCVCCCCGCCNNHNNNSCCCGGCGGCCCCGCGCGCCCCNNHN 110681

Qy    1527 acagcagcccggattcagggaagcagctggtgcactacacagcgcagccgctgttcctgc 1586
              |  |  |||   :   :    || |    :   :   | || ||        : ||
Db 110680 NNSGGCGGGCGGCNNHNNNSGGCGGGCGGCGNNHNNNSGGCGGGCGGCGCNNHNNNSGGC 110621

Qy    1587 tggaccccggctccgtggacaccgggagcaacgacctgccggtgctgtttgagctgggag 1646
            ||   |  |  :   :| |    || ||    :   :   | || |   | |      :
Db 110620 GGGCGGCGCGNNHNNNSGGCGGGCGGCGCGGNNHNNNSGGCGGGCGGCGCGGGNNHNNNS 110561

Qy    1647 agggctcctacttctccgaaggggacggcttcgccgaggaccccaccatctccctgctga 1706
            | |  |           :   :| |||      || ||   :   :  |   | ||
Db 110560 GGCGGGCGGCGCGGGGNNHNNNSGGCGGGCGGCGCGGGGCNNHNNNSGGCGGGCGGCGCG 110501

Qy    1707 caggctcggagcctcccaaagccaaggaccccactgtctcctagaggccccggaggagct 1766
              | |      :  |     ||   || |||  :   : |   | | |||       |
Db 110500 GGGCCNNHNNNSGGCGGGCGGCGCGGGGCCCCNNHNNNSGCGCGGGGCCCCNNHNNNSGGC 110441

Qy    1767 gggccagccgcccacccccacccccagtgcagggctggtcttggggaggcagggcagcct 1826
            | |    ||   :  :|  | |   |  |  |       : |   | | |||     :
Db 110440 GCGGGGCCCCCNNHNNNSCGGCGCGGGGCCCCNNHNNNSGCGGCGCGGGGCCCCNNHNNNS 110381

Qy    1827 cgcggtcttgggcactggtgggtcggccgccatagccccagtaggacaaacgggctcggg 1886
             ||||    |||| |   :     ||| ||    | |||   :   :     ||| ||||
Db 110380 GGCGGCGCGGGGCCCCNNHNNNSGGGCGGCGCGGGGCCCCNNHNNNSCGGGCGGCGCGGG 110321

Qy    1887 tctgggcagcacctctggtcaggagggtcaccctggcctgc 1927
            |     :   :    |    | ||| | |  :   : ||
Db 110320 GCCCCNNHNNNSGCGGGCGGCGCGGGGCCCCNNHNNNSGGC 110280


RESULT  10
X02974
ID   X02974 standard; DNA; 3198 BP.
AC   X02974;
DT   22-JUN-1999  (first entry)
DE   Human IL-1ra BAC contiguous DNA sequence 19.
KW   Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW   interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW   chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW   growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS   Homo sapiens.
PN   WO9906426-A1.
PD   11-FEB-1999.
PF   03-AUG-1998; U16102.
PR   02-JUL-1998; US-091650.
PR   04-AUG-1997; US-054646.
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI   Pan Y;
DR   WPI; 99-153692/13.
PT   New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT   used to inhibit inflammation and to screen for specific modulators
PS   Example 5; Figure 3; 226pp; English.
CC   X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC   containing alternatively spliced forms of human IL-1ra. Such fragments
CC   are used in the method of the invention which describes the isolation of
CC   a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC   member of the cytokine superfamily that is expected to inhibit
CC   inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC   also bind to a new receptor so could regulate other cellular processes
CC   associated with acute or chronic inflammation, e.g. asthma, chronic
CC   myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC   bowel disease. It may also induce or suppress interleukins, cytokines and
CC   growth factors. Modulators of this protein are used to treat or prevent
CC   conditions associated with abnormal levels of inflammation, or activity
CC   of IL-1 or its receptor complex.
SQ   Sequence   3198 BP;    807 A;    847 C;    681 G;    845 T;


  Query Match             2.3%;  Score 48.8;  DB 1;  Length 3198;
  Best Local Similarity   44.9%;  Pred. No. 0.32;
  Matches  182;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy     862 ccacggctcgggcccctactcggcccctccccagcctacagcagctccagcctctacgc 921
           || |  | |   ||||  | |  || || |||| |||    || |  |  ||| |  | |
Db    2789 CCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCGCCCCGNGCGGGCCGCGCCCCCCCCC 2848

Qy     922 ccctgatgctgtggccagctctggacccatcatctccgacatcaccgagctggctcctgc 981
            ||          ||| |   || |||     | ||  |  | ||   |  || || |
Db    2849 GCCCCCCCGGCCCGCCCCCCGCGGCCCCCCACCCCCCCCCCCCCCCCCCCGCGCCCCGCC 2908

Qy     982 cagccccatggcctcccccggcgggagcatagacgagaggcccctatccagcagccccct 1041
           |  ||||  |  | ||||   |    ||      ||     ||||   ||  |  |||||
Db    2909 CCCCCCCCCGCGCCCCCCACCCCCCCGCCCCCCCGCCCCCCCCCCCCCCCCCCACCCCCC 2968

Qy    1042 ggtgcgtgtcaaggaggagcccccagcccgcctcagagccccccgggtagaggaggcgag 1101
              | | | |   | | |||||   ||  |  |   ||||||         |  ||
```

```
Db   2969 ACACCCGGCCCACACGCACCCCCCACCCCGACGCCCCCGCCCCCCCCCCCCCCGCAGCCGA 3028

Qy   1102 tcccgggcgcccatcttccgtggacaccctcttgtccccgaccgccctcattgactccat 1161
Db   3029 CGCCCCCCCCCCGCCCGCCCCGCCCCGCACCCCCGACCCCCCCCGCCGCCCCGCCCCCGC 3088

Qy   1162 cctgcgggagagtgaacctgcccccgcctccgtcacagccctcacggacgccaggggcca 1221
Db   3089 CCCCCCCCCCCGGCCCCCCCCCCGCCGGCGCGGCGCCCCACCCCCCCCCCCCAGCCCCGA 3148

Qy   1222 cacggacaccgagggccggcctccctccccccgcccacctccac 1266
Db   3149 CCGCGCGCCCCCCCCACCCCCCCCCCAGCCCCCGCCCCCCGCCCC 3193


RESULT  11
V82460
ID   V82460 standard; DNA; 1791 BP.
AC   V82460;
DT   16-MAR-1999  (first entry)
DE   Triticum sp. cysteine proteinase #5 encoding DNA.
KW   Triticum; wheat seed; cysteine proteinase; gluten; baking; ds.
OS   Triticum sp.
FH   Key             Location/Qualifiers
FT   CDS             81. .1490
FT                   /*tag=  a
PN   J10327886-A.
PD   15-DEC-1998.
PF   27-MAR-1998; 098140.
PR   31-MAR-1997; JP-114946.
PA   (SHOS ) SHOWA SANGYO CO.
DR   WPI; 99-109255/10.
DR   P-PSDB; W89560.
PT   New DNA coding cysteine proteinase originating from wheat seed -
PT   useful for improving gluten for use in bakery process
PS   Claim 5; Page 22-24; 29pp; Japanese.
CC   The present sequence encodes a cysteine proteinase isolated from wheat
CC   seed (Triticum sp.). The cysteine proteinase is useful for improving
CC   gluten for use in the bakery process.
SQ   Sequence   1791 BP;    379 A;    545 C;    532 G;    333 T;


  Query Match             2.3%;  Score 48.6;  DB 1;  Length 1791;
  Best Local Similarity   46.6%;  Pred. No. 0.32;
  Matches  156;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy     34 ggccccggggctgtgtgtgcgcagcgggcggcggcgcggcccggaaggctggcgcggcga 93
Db    215 GGCCCGGGCCGTCTACGACCTCTGGCTGGCCGAGCACGGCGGCGGGTCCTACCCCAACGC 274

Qy     94 cggcgttagcccggccctcggcccctctttgcggccgctccctccgcctattccctcctt 153
Db    275 CAACTCCATCCCGGAGCGCGAGCGCCGGTTCCGGGCCTTCTGGGACAACCTCCGCTTCGT 334

Qy    154 gctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggcctt 213
Db    335 CGACGCCCACAACGCGCGCGCCGCGGCCGGCGAGGAGGGCTTCCGCCTCGCCATGAACCG 394

Qy    214 cctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggag 273
Db    395 CTTCGCCGATCTCACCAACGACGAGTTCCGCGCCGCCTACCTCGGGGTCAAGGGCCAGAG 454

Qy    274 cccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtgctgcc 333
Db    455 GGCGAGGCCCGGCCGCGTCGTCGGGGAGCGGTACCGCCACGACGGGCCGGAGGAGCTGCC 514

Qy    334 caagtacttcaagcacaacaacatggccagcttcg 368
Db    515 CGAGGCCGTCGACTGGAGGGAGAAGGGCGCCGTCG 549


RESULT  12
T85356
```

```
ID   T85356 standard; DNA; 2004 BP.
AC   T85356;
DT   09-DEC-1997  (first entry)
DE   Nephila clavipes spider silk protein 2 Kb DNA sequence.
KW   High strength film; fibre; woven article; parachutes; sails;
KW   absorber; body armour; heavy metal; biological weapon; chemical;
KW   flavour; fragrance; Nephila clavipes; ss.
OS   Nephila clavipes.
FH   Key             Location/Qualifiers
FT   cds             40. .1980
FT                   /*tag=  a
FT                   /product= Silk_protein
PN   WO9708315-A1.
PD   06-MAR-1997.
PF   22-AUG-1996; U13767.
PR   22-AUG-1995; US-517694.
PA   (BASE/) BASEL R M.
PA   (ELIO/) ELION G R.
PI   Basel RM,  Elion GR;
DR   WPI; 97-179272/16.
DR   P-PSDB; W27178.
PT   New opt. multimerised DNA sequences encoding spider silk protein -
PT   contg. both repetitive and non-repetitive sequences, useful for
PT   making high strength films, fibres, woven articles etc.
PS   Claim 14; Fig 1; 57pp; English.
CC   A process has been developed for the production of a DNA fragment
CC   encoding silk protein. The process involves: (a) selecting target DNA,
CC   from a silk-producing spider, that contains many repetitive and non-
CC   repetitive regions; (b) selecting a single-stranded DNA primer of at
CC   least 10 nucleotides with a sequence that is complementary to a region
CC   of the target; (c) repetitively combining the primer with melted target
CC   DNA, incubating the mixture with nucleotides and a DNA polymerase with
CC   proofreading activity to produce a DNA fragment which is complementary
CC   to the target and is at least 2 kb long. The present sequence
CC   represents a 2 Kb DNA sequence which encodes the spider silk protein
CC   from Nephila clavipes. The DNA fragment can be used to make fibres,
CC   films, woven articles, e.g. for use in parachutes, sails, body armour,
CC   and absorbers (e.g. of heavy metals, biological weapons, DNA, chemicals,
CC   flavours and fragrances). The high molecular weight (90-250 kD) of
CC   spider silk proteins can be produced on a commercial scale (at over
CC   2 g/l cell mass). It has better tensile strength and elasticity than
CC   silkworm silk. Inclusion of both repetitive and non-repetitive regions
CC   ensures isolation of stable clones.
SQ   Sequence   2004 BP;    481 A;    386 C;    791 G;    346 T;


  Query Match             2.3%;  Score 48.6;  DB 1;  Length 2004;
  Best Local Similarity   45.4%;  Pred. No. 0.33;
  Matches  214;  Conservative    0;  Mismatches  254;  Indels    3;  Gaps    1;

Qy    317 gccaaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcag 376
Db    266 GACAAGGAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCAGCCGCTGCAGCAG 325

Qy    377 ctcaacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagcca 436
Db    326 CTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCTGGACGAGGTG 385

Qy    437 gagagagacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctcctt 496
Db    386 GACAAGGTGCAGGCGCAGCCGCAGCAGCAGCCGGAGGTGCTGGACAAGGAGGATACGGTG 445

Qy    497 gagaacatcaagaggaaagtgaccagtgtg---tccaccctgaagagtgaagacataaag 553
Db    446 GACAAGGTGCCGGACAAGGAGGCTATGGAGGACTTGGAAGTCAAGGTGCTGGACGAGGAG 505

Qy    554 atccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcag 613
Db    506 GATTAGGTGGACAAGGTGCAGGTGCAGCAGCAGCAGCAGCTGGAGGTGCCGGACAGG 565

Qy    614 gagtgcatggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggag 673
Db    566 GAGGATTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCAGCCGCTGCAGCAGCTGGTG 625
```

```
Qy    674 gtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcag 733
          ||| |   |     | | | | | | |  | ||  || ||     || || |   ||||
Db    626 GTGCCGGACAAGGAGGATATGGAGGTCTCGGAAGCCAAGGTGCAGGACGAGGTGGATCAG 685

Qy    734 ttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagagaaag 784
          |   |     | |||||||    | | |   |||| ||||  |   |||
Db    686 GTGGACAAGGGGCAGGTGCAGCAGCAGCAGCAGCTGGAGGTGCCGGACAAG 736


RESULT  13
Q14183
ID   Q14183 standard; cDNA; 2338 BP.
AC   Q14183;
DT   15-JAN-1992  (first entry)
DE   N.clavipes dragline silk protein coding sequence.
KW   protein superfibre; major ampullate silk; orb web spider; ss.
OS   Nephilia clavipes.
FH   Key             Location/Qualifiers
FT   cds             1. .2157
FT                   /*tag=  a
FT                   /product= silk protein 1
PN   EP-452925-A.
PD   23-OCT-1991.
PF   18-APR-1991; 106217.
PR   20-APR-1990; US-511792.
PA   (UYWY-) UNIV OF WYOMING.
PI   Lewis RV, Xu M, Hinman M;
DR   WPI; 91-312199/43.
DR   P-PSDB; R14308.
PT   DNA encoding spider silk protein-1 and 2 and variants - isolated
PT   from Nephila clavipes, for prodn. of spider silk protein and
PT   fibres having desired characteristics
PS   Claim 4; Page 23; 48pp; English.
CC   A N.clavipes major ampullate gland cDNA library was screened with
CC   probes based on peptide fragments of the purified spider silk
CC   protein (see Q14185). Positive plaques were identified and the
CC   spider silk protein 1 coding sequence was determined. See also
CC   Q14184.
SQ   Sequence   2338 BP;    566 A;    433 C;    916 G;    423 T;


  Query Match             2.2%;  Score 46.8;  DB 1;  Length 2338;
  Best Local Similarity   45.7%;  Pred. No. 0.76;
  Matches  202;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    1;

Qy    317 gccaaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcag 376
          | ||||||||    |    |  |  |   | ||||  |  |||          ||| ||||
Db    680 GACAAGGAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCATCCGCTGCAGCAG 739

Qy    377 ctcaacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagcca 436
          ||     ||   | |     | |  |||    | | ||       |  |||| | ||
Db    740 CTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCTGGACGAGGTG 799

Qy    437 gagagagacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctcctt 496
          ||||  |   |  || ||   ||||| |          |||  | | |||||  |     |
Db    800 GAGAAGGTGCAGGCGCAGCCGCAGCAGCAGCCGGAGGTGCTGGACAAGGAGGATACGGTG 859

Qy    497 gagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagaca---taaag 553
          |     |  | | |     |  ||  | | | |    | |    |||| ||   |   |
Db    860 GTCTTGGTGGACAAGGTGCCGGACAAGGAGGCTATGGAGGACTTGGAAGCCAAGGTGCTG 919

Qy    554 atccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcag 613
            |    ||||    ||   |||| ||| |  | |   |||||||  |   |   |    ||
Db    920 GACGAGGAGGATTAGGTGGACAAGGTGCAGGTGCAGCAGCAGCTGGAGGTGCCGGGCAAG 979

Qy    614 gagtgcatggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggag 673
          |||  |  ||    |||| | |||| || | |||  ||| |   |   | |  || || |
Db    980 GAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCAGCCGCTGCAGCAGCTGGTG 1039

Qy    674 gtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcag 733
          ||| |   |     | | | | | | |  | ||  || ||     || || |   | ||
Db   1040 GTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCAGGACGAGGTGGATTAG 1099

Qy    734 ttcctgatctcactggtgcagt 755
          |   |     | ||||||||
Db   1100 GTGGACAAGGGGCAGGTGCAGT 1121


RESULT  14
V23249
ID   V23249 standard; cDNA; 2338 BP.
AC   V23249;
DT   06-JUL-1998  (first entry)
DE   Nephila clavipes spider silk protein encoding cDNA.
KW   Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;
KW   cocoon; tensile strength; elasticity; ss.
OS   Nephila clavipes.
FH   Key             Location/Qualifiers
FT   CDS             1. .2157
FT                   /*tag=  a
FT                   /product= "spider silk protein 1"
PN   US5728810-A.
PD   17-MAR-1998.
PF   19-APR-1995; 425069.
PR   15-APR-1991; US-684819.
PR   20-APR-1990; US-511792.
PR   04-OCT-1994; US-317844.
PR   19-APR-1995; US-425069.
PA   (UYWY-) UNIV WYOMING.
PI   Hinman MB, Lewis RV, Xu M;
DR   WPI; 98-270437/24.
DR   P-PSDB; W53346.
PT   Recombinant spider silk proteins - useful for making fibres
PS   Example 3; Column 25-30; 68pp; English.
CC   The present sequence encodes a spider silk protein from the present
CC   invention. Spider silk proteins, and peptide fragments within the
CC   proteins, can be produced and purified independently and can then be
CC   mixed and made into fibres that have higher tensile strengths and
CC   elasticity than naturally occurring fibres. The fibres can be used in
CC   mixed composites. The invention allows the two naturally occurring
CC   Nephila clavipes silk proteins to be produced independently so that
CC   they can later be combined to form silk fibres of high tensile strength
CC   and elasticity.
SQ   Sequence   2338 BP;    564 A;    432 C;    918 G;    424 T;


  Query Match             2.2%;  Score 46.8;  DB 1;  Length 2338;
  Best Local Similarity   45.7%;  Pred. No. 0.76;
  Matches  202;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    1;

Qy    317 gccaaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcag 376
          | ||||||||    |    |  |  |   | ||||  |  |||          ||| ||||
Db    680 GACAAGGAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCATCCGCTGCAGCAG 739

Qy    377 ctcaacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagcca 436
          ||     ||   | |     | |  |||    | | ||       |  |||| | ||
Db    740 CTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCTGGACGAGGTG 799

Qy    437 gagagagacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctcctt 496
          ||||  |   |  || ||   ||||| |          |||  | | |||||  |     |
Db    800 GAGAAGGTGCAGGCGCAGCCGCAGCAGCAGCCGGAGGTGCTGGACAAGGAGGATACGGTG 859

Qy    497 gagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagaca---taaag 553
          |     |  | | |     |  ||  | | | |    | |    |||| ||   |   |
Db    860 GTCTTGGTGGACAAGGTGCCGGACAAGGAGGCTATGGAGGACTTGGAAGCCAAGGTGCTG 919

Qy    554 atccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcag 613
            |    ||||    ||   |||| ||| |  | |   |||||||  |   |   |    ||
Db    920 GACGAGGAGGATTAGGTGGACAAGGTGCAGGTGCAGCAGCAGCTGGAGGTGCCGGGCAAG 979

Qy    614 gagtgcatggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggag 673
          |||  |  ||    |||| | |||| || | |||  ||| |   |   | |  || || |
Db    980 GAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCAGCCGCTGCAGCAGCTGGTG 1039

Qy    674 gtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcag 733
```

```
Db   1040 GTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCAGGACGAGGTGGATTAG 1099

Qy    734 ttcctgatctcactggtgcagt 755
Db   1100 GTGGACAAGGGGCAGGTGCAGT 1121
```

RESULT 15
Q93037
ID Q93037 standard; cDNA; 2517 BP.
AC Q93037;
DT 07-DEC-1995 (first entry)
DE Human E2F-1 complete cDNA.
KW E2F-1; cell culture; serum-free; protein-free; culture medium;
KW cell cycle; transcription factor; pRB-binding protein;
KW DNA binding protein; CHO-K1; mammal cell culture; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 136. .1449
FT /*tag= a
FT /product= E2F-1
PN EP-666312-A.
PD 09-AUG-1995.
PF 05-FEB-1995; 101534.
PR 08-FEB-1994; CH-000363.
PA (BAIL/) BAILEY J E.
PA (EPPE/) EPPENBERGER H M.
PA (RENN/) RENNER W A.
PI Bailey JE, Eppenberger HM, Renner WA;
DR WPI; 95-270494/36.
PT Animal cell cultures able to grow in serum and protein-free medium -
PT have increased content of cell cycle regulatory protein, used for
PT prodn. of diagnostic and therapeutic agents, vaccines, etc
PS Disclosure; Page 16-18; 19pp; English.
CC cDNA encoding human transcription factor E2F-1 was isolated from
CC a HeLa cDNA library. The cDNA was expressed in CHO K1 cells,
CC allowing growth of the cells in serum- and protein-free medium.
SQ Sequence 2517 BP; 454 A; 784 C; 781 G; 498 T;

Query Match 2.2%; Score 46.4; DB 1; Length 2517;
Best Local Similarity 49.2%; Pred. No. 0.92;
Matches 122; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy      38 ccggggctgtgtgtgcgcagcgggcggcggcgcggcccggaaggctggcgcggcgacggc 97
Db       7 CCGTGGCCGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGCGGGATCGAGCCCTCGCC 66

Qy      98 gttagcccggccctcggcccctctttgcggccgctccctccgcctattccctccttgctc 157
Db      67 GAGGCCTGCCGCCATGGGCCCGCGCCGCCGCCGCCGCCTGTCACCCGGGCCGCGCGGGCC 126

Qy     158 gagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctg 217
Db     127 GTGAGCGTCATGGCCTTGGCCGGGGCCCCTGCGGGCGGCCCATGCGCGCCGGCGCTGGAG 186

Qy     218 accaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccg 277
Db     187 GCCCTGCTCGGGGCCGGCGCGCTGCGGCTGCTCGACTCCTCGCAGATCGTCATCATCTCC 246

Qy     278 agcgggaa 285
Db     247 GCCGCGCA 254
```

Search completed: March 6, 2000, 21:20:27
Job time: 3622 sec

```
OM of: US-09-304-121-2 to: Issued_Patents_NA:*   out_format : pfs

Date: Mar 7, 2000 12:19 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09304121/runat_06032000_151317_15723/app_query.fasta.1
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09304121 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-304-121-2
Query length: 529
Database: Issued_Patents_NA:*
Database sequences: 214294
Database length: 59861574
Search time (sec): 27.560000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-178-477B-31 + 2729.00 4287.86  2.1e-231  2156  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-178-477B-42 +  705.50 1080.81  9.1e-53   2781  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-130-114-2 +  165.00  229.56  2.4e-05   1931  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-844-298B-1 +  155.00  203.94  0.0006    4237  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-764-233A-1 +  155.00  173.36  0.0321   49377  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-258-261B-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-456-837-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-342-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-646A-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-458-076A-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-764-233A-4 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-457-335A-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-729-214-6 +  146.50  166.54  0.0768   28958  !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-864-038A-1 -  145.00  196.19  0.0017    2214  !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-864-038A-2 -  145.00  191.11  0.0033    3331  !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-864-038A-4 -  145.00  191.11  0.0033    3331  !
/cgn2_6/ptodata/1/ina/backfiles1.seq:5196516-7 +  140.50  179.18  0.0152    4897  ! Pate
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-945-283-1 -  139.00  170.03  0.0491    8438  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-440-861-1 +  137.50  191.65  0.0031    1229  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-433-854-1 +  137.50  191.65  0.0031    1229  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-174-745A-1 +  137.50  191.65  0.0031    1229  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-195-947-1 +  137.50  191.65  0.0031    1229  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-433-885-1 +  137.50  191.65  0.0031    1229  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-433-908B-1 +  137.50  191.65  0.0031    1229  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-853-913-1 -  136.50  162.50  0.1289   11236  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-653-091A-11 +  134.50  177.81  0.0181    2549  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-485-068-11 +  134.50  177.81  0.0181    2549  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-481-598-11 +  134.50  177.81  0.0181    2549  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-483-353-11 +  134.50  177.81  0.0181    2549  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-473-702-11 +  134.50  177.81  0.0181    2549  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-290-731C-3 +  133.00  166.43  0.0779    5253  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-452-655B-1 +  133.00  160.39  0.1690    8532  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-741-940-1 +  133.00  158.91  0.2043    9606  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-289-548A-1 +  133.00  158.91  0.2043    9606  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-452-654-1 +  133.00  158.91  0.2043    9606  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-370-235A-1 +  133.00  158.91  0.2043    9606  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-885-970A-25 +  132.50  177.77  0.0182    1984  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-298-687A-25 +  132.50  177.76  0.0182    1985  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-298-829-25 +  132.50  177.76  0.0182    1985  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-993-118-1 +  132.50  168.94  0.0565    4031  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-993-118-9 +  132.50  168.35  0.0609    4225  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-1 +  132.00  166.68  0.0754    4534
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-244-646-14 +  131.00  182.56  0.0098    1116
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-592-936B-20 +  131.00  170.59  0.0457    2917
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-111-573-20 +  131.00  170.59  0.0457    2917

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-178-477B-31

seq_documentation_block:
; Sequence 31, Application US/08178477B
; Patent No. 5756343
;  GENERAL INFORMATION:
;    APPLICANT:  WU, CARL; CLOS, JOACHIM;
;    APPLICANT:  WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
;    TITLE OF INVENTION:  CELL STRESS
;    TITLE OF INVENTION:  TRANSCRIPTIONAL FACTORS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/178,477B
;      FILING DATE:  07-JAN-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/617,910
;      FILING DATE:  26-NOV-1990
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4103US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2156
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  (DNA) genomic
US-08-178-477B-31


alignment_scores:
            Quality: 2729.00              Length:    529
              Ratio:  5.159                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-304-121-2 x US-08-178-477B-31   ..

 Align seg 1/1  to: US-08-178-477B-31  from: 1  to: 2156

       1 MetAspLeuProValGlyProGlyAlaAlaGlyProSerAsnValProAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 ATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC 210

      17 aPheLeuThrLysLeuTrpThrLeuValSerAspProAspThrAspAlaL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGC 260

      34 euIleCysTrpSerProSerGlyAsnSerPheHisValPheAspGlnGly 50
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 261 TCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTTCGACCAGGGC 310

  51 GlnPheAlaLysGluValLeuProLysTyrPheLysHisAsnAsnMetAl 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 311 CAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

  67 aSerPheValArgGlnLeuAsnMetTyrGlyPheArgLysValValHisI 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACA 410

  84 leGluGlnGlyGlyLeuValLysProGluArgAspAspThrGluPheGln 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 411 TCGAGCAGGGCGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAG 460

 101 HisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysAr 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 461 CACCCATGCTTCCTGCGTGGCCAGGAGCAGCTCCTTGAGAACATCAAGAG 510

 117 gLysValThrSerValSerThrLeuLysSerGluAspIleLysIleArgG 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 511 GAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC 560

 134 lnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLysGlyLys 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 561 AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAG 610

 151 GlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAl 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 611 CAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

 167 aLeuTrpArgGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnL 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGA 710

 184 ysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsn 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 711 AAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAAC 760

 201 ArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGl 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 761 CGGATCCTGGGGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTGG 810

 217 ySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisV 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 811 CTCAGCACATTCCATGCCCAAGTATAGCCGGCAGTTCTCCCTGGAGCACG 860

 234 alHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSer 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 861 TCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTCC 910

 251 SerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAs 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 911 AGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

 267 pIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerI 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCA 1010

 284 leAspGluArgProLeuSerSerSerProLeuValArgValLysGluGlu 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1011 TAGACGAGAGGCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAG 1060

 301 ProProSerProProGlnSerProArgValGluGluAlaSerProGlyAr 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1061 CCCCCCAGCCCGCCTCAGAGCCCCCGGGTAGAGGAGGCGAGTCCCGGGCG 1110

 317 gProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerI 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1111 CCCATCTTCCGTGGACACCCTCTTGTCCCCGACCGCCCTCATTGACTCCA 1160

 334 leLeuArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAsp 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1161 TCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGCCCTCACGGAC 1210

 351 AlaArgGlyHisThrAspThrGluGlyArgProProSerProProProTh 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1211 GCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCGCCCAC 1260

 367 rSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGluL 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGC 1310

 384 euSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThr 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1311 TCAGTGACCACTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACC 1360

 401 MetLeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1361 ATGCTGAGCAGCCACGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCT 1410

 417 uPheSerProSerValThrValProAspMetSerLeuProAspLeuAspS 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1411 GTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACA 1460

 434 erSerLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProArg 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1461 GCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGG 1510

 451 ProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValHi 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1511 CCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

 467 sTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThrG 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCG 1610

 484 lySerAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPhe 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1611 GGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTC 1660

 501 SerGluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1661 TCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGG 1710

 517 ySerGluProProLysAlaLysAspProThrValSer 529
     |||||||||||||||||||||||||||||||||||
1711 CTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC 1747
```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-178-477B-42

seq_documentation_block:
; Sequence 42, Application US/08178477B
; Patent No. 5756343
;  GENERAL INFORMATION:
;    APPLICANT:  WU, CARL; CLOS, JOACHIM;
;    APPLICANT:  WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
;    TITLE OF INVENTION:  CELL STRESS
;    TITLE OF INVENTION:  TRANSCRIPTIONAL FACTORS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/178,477B
;      FILING DATE:  07-JAN-1994
;      CLASSIFICATION:  530
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/617,910
;       FILING DATE:  26-NOV-1990
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CAROL M. GRUPPI
;       REGISTRATION NUMBER:  37,341
;       REFERENCE/DOCKET NUMBER:  2026-4103US1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;       TELEX:  421792
;   INFORMATION FOR SEQ ID NO:  42:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2781
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-178-477B-42


alignment_scores:
            Quality:  705.50              Length:    665
              Ratio:  2.016                 Gaps:     23
 Percent Similarity:  52.632   Percent Identity: 30.526

alignment_block:
 US-09-304-121-2 x US-08-178-477B-42   ..

 Align seq 1/1  to: US-08-178-477B-42  from: 1  to: 2781

       6 GlyProGlyAlaAlaGlyProSerAsnValProAlaPheLeuThrLysLe 22
         |||   :::||||||      |||:::|||||||||||||||||:::|||||
     337 GGAGACGCCGCGGCCATCGGAAGCGGGGTGCCGGCCTTTTTGGCCAAATT 386

      22 uTrpThrLeuValSerAspProAspThrAspAlaLeuIleCysTrpSerP 39
         ||||   ||||||:::|||   ||||||:::   ||||||||||||:::
     387 GTGGCGCCTGGTGGACGATGCCGATACCAATCGCTTGATTTGCTGGACCA 436

      39 roSerGlyAsnSerPheHisValPheAspGlnGlyGlnPheAlaLysGlu 55
           :::|||:::||||||   :::   :::|||:::|||||||||||||||
     437 AGGATGGCCAAAGTTTCGTTATTCAAAATCAAGCGCAATTTGCCAAGGAA 486

      56 ValLeuProLysTyrPheLysHisAsnAsnMetAlaSerPheValArgGl 72
         :::||||||      :::|||||||||||||||||||||||||:::|||||
     487 CTATTGCCACTAAACTACAAGCACAACAACATGGCCAGTTTCATAAGGCA 536

      72 nLeuAsnMetTyrGlyPheArgLysValValHisIleGluGlnGlyGlyL 89
         |||||||||||||||||||:::|||:::::   |||:::::|||||||
     537 ATTGAATATGTATGGATTCCACAAGATCACCTCTATTGACAATGGCGGAC 586

      89 euValLysProGluArgAspAspThrGluPheGlnHisProCysPheLeu 105
         ||   :::   :::||||||:::   ||||||:::||||||   |||
     587 TA...CGTTTTGATCGCGACGAGATTGAATTTTCGCACCCATTTTTTAAG 633

     106 ArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSerVa 122
         |||::::::      ||||||:::::::||||||||||||         ::
     634 CGCAACTCGCCTTTTCTACTTGACCAAATCAAAAGGAAA.....AT 674

     122 lSerThrLeuLysSerGluAspIleLys......IleArgGlnAspServ 137
         :|||:::   |||:::   |||   |||      ::::::   :::::::
     675 ATCGAACAACAAAAATGGTGACGACAAAGGTGTCCTGAAGCCGGAGGCCA 724

     137 alThrLysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCys 153
         :::::|||:::||||||||||||:::::::|||:::|||:::|||:::
     725 TGTCGAAGATTCTCACCGATGTGAAAGTCATGCGGGGTCGTCAGGACAAT 774

     154 MetAspSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpAr 170
         :::||||||::::::   |||||||||:::||||||||:::||||||||
     775 CTGGATTCGCGCTTCTCCGCCATGAAACAGGAGAACGAAGTGCTGTGGCG 824

     170 gGluValAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValA 187
         ||||:::||||||||||||||||||||||||:::||||||::::::||||
     825 CGAGATAGCCAGCCTGCGCCAAAAGCACGCTAAGCAGCAACAAATAGTCA 874

     187 snLysLeuIleGlnPheLeuIleSerLeuValGlnSerAsnArg...Ile 202
         ||||||||||||||||||||||||:::::||||||   :::|||   :::
     875 ACAAACTGATCCAGTTCCTCATTACCATTGTGCAACCGTCGCGCAACATG 924

     203 LeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySerAl 219
         ||||||||||||   :::   ||||||:::|||::::::
     925 TCTGGCGTGAAACGCCATGTGCAGCTGATGATCAACAATACG..... 966

     219 aHisSerMetProLysTyrSerArg.........GlnPheSerLeuGluH 233
                 |||:::   :::|||           :::   |||
     967 .........CCGGAAATTGATCGTGCACGGACCACCAGTGAGACCGAGA 1006

     233 isValHisGlySerGlyPro..*.............................. 239
               |||:::||||||
    1007 GCGAGAGTGGCGGCGGACCGGTTATCCACGAGCTTAGGGAGGAGCTTCTT 1056

     240 ......TyrSerAlaProSerProAla...TyrSerSerSerSerLeuTy 253
                       ||||||||||||   |||:::::::::|||   ||
    1057 GATGAGGTGATGAATCCATCACCGGCTGGCTACACCGCAGCCTCACATTA 1106

     253 rAlaProAspAlaValAla.............................. 259
         |      ::::::|||:::
    1107 TGACCAAGAGAGCGTCTCTCCGCCTGCCGTTGAGCGTCCGCGATCTAACA 1156

     259 ..... 259

    1157 TGAGCATTAGCTCGCACAACGTCGATTATTCGAATCAGAGTGTGGAGGAC 1206

     260 ............SerSerGlyProIleIleSerAspIleThrGluLeuAl 272
                     ::::::|||         :::::|||         ::
    1207 TTGCTGCTCCAGGGAAATGGAACCGCTGGCGGTAATATTCTAGTAGGCGG 1256

     272 aProAlaSerProMetAlaSerProGlyGlySerIleAspGluArgPro. 288
         :   ||||||||||||||::         |||:::::::   |||
    1257 AGCCGCTTCTCCCATGGCCCAA.........AGTGTGAGTCAATCGCCGG 1297

     289 ...................LeuSerSerSerProLeuValArgValLys 298
                             :::::::::::::|||      :::|||:::
    1298 CCCAACATGATGTCTACACAGTCACCGAGGCGCCCGATTCTCATGTCCAG 1347

     299 GluGluProProSerProPro........................... 305
         |||   |||   ||||||||
    1348 GAGGTGCCAAACAGTCCGCCTTATTACGAGGAGCAGAATGTGCTTACCAC 1397

     306 ......................GlnSerProArg................ 309
                                 |||:::   |||
    1398 GCCCATGGTGCGGGAGCAGGAGCAGCAGAAGCGTCAGCAGCTTAAGGAGA 1447

     309 ..... 309

    1448 ACAACAAGCTACGACGACAGGCAGGAGATGTTATACTTGATGCTGGAGAT 1497

     310 ......ValGluGluAlaSerPro...GlyArgProSerSerValAspTh 323
               |||::::::::::||||||   ::::::   :::|||::::::
    1498 ATTCTCGTAGATAGTTCGTCGCCCAAGGCGCAACGGACAAGCATCCAGCA 1547

     323 rLeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluP 340
               :::|||   ::::::::::::   ::::::   :::||||||||
    1548 TAGTACGCAACCTGATGTGATGGTCCAGCCAATGATTATAAAGTCTGAGC 1597

     340 roAlaProAlaSer......ValThrAlaLeuThrAspAlaArgGly... 353
         ||      :::|||      :::      :::|||   |||:::
    1598 CGGAGAACAGTTCCGGACTGATGGATCTAATGACTCCCGCGAACGATCTG 1647

     353 ..... 353

    1648 TACAGTGTCAACTTCATCAGTGAGGATATGCCGACGGATATTTTTGAAGA 1697

     353 ..... 353
```

```
    1698 CGCTCTGCTTCCCGACGGCGTGGAAGAGGCAGCCAAACTGGACCAGCAGC 1747

     354 .............HisThrAspThrGluGlyArgProProSer...... 363
                      |||   ::::::|||:::      |||
    1748 AAAAATTTGGGCAATCGACAGTGAGCAGCGGCAAGTTTGCCAGCAACTTC 1797

     364 ...ProProProThrSerThr.......................... 369
              |||   :::||||||
    1798 GATGTGCCCACCAACAGTACGCTGCTGGATGCCAATCAGGCCTCGACATC 1847

     370 ............................ProGluLysCysLeuSerValA 377
                                      |||:::   ::::::||||
    1848 GAAGGCAGCGGCCAAGGCGCAAGCATCTGAGGAAGAGGGCATGGCTGTGG 1897

     377 la.......................................... 377
         ||
    1898 CCAAATACAGTGGCGCTGAGAACGGAAACAACCGCGATACCAACAACAGT 1947

     378 ............CysLeuAspLysAsnGluLeuSerAspHisLeuAspAl 390
                     :::   :::||||||      ||||||:::::
    1948 CAACTCCTCAGGATGGCCTCAGTTGACGAACTCCACGGGCACTTGGAAAG 1997

     390 aMetAspSerAsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyP 407
         :|||:::::::::|||:::::::|||:::   :::|||   :::   |||
    1998 CATGCAAGATGAGTTGGAAACACTGAAGGATCTGCTGCGCGGCGATGGGG 2047

     407 heSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerValThr 423
           :::::::|||   :::   |||:::   ||||||:::   |||
    2048 TGGCCATTGATCAGAACATGCTCATGGGTCTGTTTAACGACTCTGATCTA 2097

     424 ValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGl 440
         :::   :::   :::|||         :::::::|||:::::::|||
    2098 ATGGACAACTATGGCCTATCGTTTCCCAATGACAGCATAAGCAGT..... 2142

     440 nGluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnS 457
                                                 |||::::
    2143 .........................................GAAAAGA 2149

     457 erSerProAspSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeu 473
         :::::|||   ||||||:::::::|||:::   |||      ||||||:::
    2150 AAGCACCC...AGTGGCTCTGAACTGATTTCCTAT......CAGCCCATG 2190

     474 Phe...LeuLeuAspProGlySerValAspThrGlySerAsnAsp 487
         :::   |||   |||      :::::::|||   |||:::||||||
    2191 TATGATCTGTCCGACATTTTGGACACGGACGATGGCAACAATGAC 2235

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-130-114-2

seq_documentation_block:
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert A.
;  APPLICANT: Damaj, Bassam B.
;  APPLICANT: Robbins, Alan K.
;  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;  TITLE OF INVENTION: From Multiple Transfected Episomes
;  FILE REFERENCE: 0867/1D903US1
;  CURRENT APPLICATION NUMBER: US/09/130,114
;  CURRENT FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
US-09-130-114-2


alignment_scores:
           Quality:  165.00            Length:    366
             Ratio:  0.829              Gaps:     16
 Percent Similarity:  54.372   Percent Identity: 24.863

alignment_block:
 US-09-304-121-2 x US-09-130-114-2   ..

 Align seg 1/1  to: US-09-130-114-2  from: 1  to: 1931

    189 LeuIleGlnPheLeuIleSerLeuValGlnSerAsnArgIleLeuGlyVa 205
        |||:::   :::|||:::   |||:::            :::||||||::
    122 CTCCTGCCCCTTCTCCTGCTCCTGCTCCTCCGCCTTCTGGTCCTCGGGGC 171

    205 lLysArg...LysIleProLeuMet.LeuAsnAspSerGly......... 217
        :::::||      :::|||:::::: ::::::   :::|||
    172 CCGCCGAGTCCTAGTCCCGGTTCTGTATCTCTACCACAGGCCTCTGGGGT 221

    218 .............................................Se 218
                                                     ||
    222 TTTTGCAGGTTCAACGTAACCGACGTTTCCCTGGGTGCCACCTTGTCCTC 271

    218 rAlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisValH 235
        |:::   |||   |||:::   |||
    272 GTCCTCGTCCTCGCCCTCCCCGTCC..................... 296

    235 isGlySerGlyProTyrSerAlaProSerProAlaTyrSerSerSerSer 251
          :::|||:::|||   |||:::   ||||||:::   ||||||||||||
    297 ..TCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC...TCGTCCTCCTCC 341

    252 LeuTyrAlaProAspAlaValAlaSerSerGlyProIleIleSerAspIl 268
              :::|||::::::   :::|||   :::         |||
    342 CCGTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTC 391

    268 eThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerIleA 285
         ::::::   :::|||:::||||||   :::||||||::::::|||
    392 CTCGTCCTCCTCCCCGTCCTCCCCG...TCCTCCCCGTCCTCGTCC.... 434

    285 spGluArgProLeuSerSerSerProLeuValArgValLysGluGluPro 301
          :::   |||   |||||||||
    435 ..TCCTCCCCGTCCTCGTCCTCC.....................TCC 458

    302 ProSerProProGlnSerProArgValGluGluAlaSerProGlyArgPr 318
        ||||||   |||:::|||            ::::::||||||:::   ||
    459 CCGTCCTCCCCGTCCTCG............TCCTCCTCCCCGTCCTCCCC 496

    318 oSerSerValAspThrLeuLeuSerProThrAlaLeuIleAspSerIleL 335
        |||||||   ::::::      ||||||:::::      :::
    497 GTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCT 546

    335 euArgGluSerGluProAlaProAlaSerValThrAlaLeuThrAspAla 351
              :::|||:::   :::|||:::|||   ::::::   :::::::::
    547 CCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCC 596

    352 ArgGlyHisThrAspThrGluGlyArgProProSerProProProThrSe 368
          :::   ::::::::::::            ||||||   |||   :::||
    597 TCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTC 646

    368 rThrProGluLysCysLeuSerValAlaCysLeuAspLysAsnGluLeuS 385
        |:::|||::::::      |||               ::::::      |
    647 CTCCCCGTCCTCGTCCTCCTCC............CCGTCCTCCCCGTCCT 684

    385 erAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMet 401
        ||         ::::::   :::|||:::       :::   ::::::
    685 CCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCG 734

    402 LeuSerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPh 418
          ||||||      |||   ::::::|||
    735 TCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCC.............. 767

    418 eSerProSerValThrValProAspMetSerLeuProAspLeuAspSerS 435
         |||||||||         |||:::   |||   |||:::      ||||
    768 .TCCCCGTCC......TCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCT 810

    435 erLeuAlaSerIleGlnGluLeuLeuSerProGlnGluProProArgPro 451
        ||   :::|||   ::::::      |||   ::::::|||      |||
    811 CGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCG 860
```

```
     452 ProGluAlaGluAsnSerSerProAspSerGlyLysGlnLeuValHisTy 468
     861 TCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTC 910

     468 rThrAlaGlnProLeuPheLeuLeuAspProGlySerValAspThrGlyS 485
     911 CTCGTCCTCCCCG.....TCCTCCCCGTCCTCGTCCTCCTCCCCGT 951

     485 erAsnAspLeuProValLeuPheGluLeuGlyGluGlySerTyrPheSer 501
     952 CCTCGTCCTCCCCG.............TCCTCGTCCTCCACCTCCGGC 986

     502 GluGlyAspGlyPheAlaGluAspProThrIleSerLeuLeuThrGly.. 517
     987 CCCAGCTCCTCCGTCACCTCCGGCCCCAGCTCCTCCATCACCTCCGGCCC 1036

     518 .......Ser.GluProProLysAlaLysAspProThrValSer 529
    1037 CAGCTCCTCCATCACCTCCGGCGGCCCCATCTCCTGCACTTTCT 1080

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-844-298B-1

seq_documentation_block:
; Sequence 1, Application US/07844298B
; Patent No. 5272256
;  GENERAL INFORMATION:
;    APPLICANT:  Donald Benit Bloch
;    TITLE OF INVENTION:  NOVEL NUCLEAR AUTOANTIGEN
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/844,298B
;      FILING DATE:  19920202
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  Reg. No. 5272256 30,162
;      REFERENCE/DOCKET NUMBER:  00786/118001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4237
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-844-298B-1


alignment_scores:
            Quality:  155.00            Length:    450
              Ratio:  0.735              Gaps:     21
 Percent Similarity:  46.889   Percent Identity: 24.000

alignment_block:
 US-09-304-121-2 x US-07-844-298B-1   ..

Align seg 1/1  to: US-07-844-298B-1  from: 1  to: 4237

    105 LeuArgGlyGlnGluGlnLeuLeuGluAsnIleLysArgLysValThrSe 121
   1022 CTCTCAGCAGAATACCTGATTCTCAGCGATGTGCAACGGAAGGTCCTCTA 1071

    121 rValSerThrLeu......LysSerGluAspIleLysIleArgGlnAspS 136
   1072 TGTGATGGAGCTGCTGCAAAACCAGGAGGAGGGCCACGCCTGCTTCAGCT 1121

    136 erValThrLys...LeuLeuThrAspValGlnLeuMetLysGlyLysGln 151
   1122 CCATCTCGGAGTTCCTGCTCACCCACCCTGTGCTGAGCTTTGGTATCCAG 1171

    152 ..............GluCysMetAspSerLysLeuLeuAlaMetLysHi 163
   1172 GTTGTGAGTCGCTGCCGGCTACGGCACACTGAGGTGCTGCCTGCCGAAGA 1221

    163 sGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisA 180
   1222 GGAAAATGACAGCCTGGGT.........GCTGATGGTACCCATGGAGCCG 1262

    180 laGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleSerLeu 196
   1263 GTGCCATGGAGTCTGCGGCCGGTGTGCTCATCAAG......CTCTTTTGT 1306

    197 ValGlnSerAsnArgIleLeuGlyValLysArgLysIleProLeuMetLe 213
   1307 GTGCATACTAAGGCACTGCAAGATGTGCAGATCCGCTTCCAGCCACAGCT 1356

    213 uAsn.....AspSerGlySerAlaHis..... 220
   1357 GAACCCTGATGTGGTGGCCCCACTGGGGACCCACACTGCCCACGAGGACT 1406

    221 ..............SerMetProLysTyrSerArgGlnPheSerLeuGlu 232
   1407 TCACATTTGGAGAGTCTCGGCCCGAACTGGGCTCTGAGGGCCTGGGGTCA 1456

    233 HisValHisGlySerGlyPro.....TyrSerAl 242
   1457 GCCGCTCACGGCTCCCAGCCTGACCTCCGACGAATCGTGGAGCTGCCTGC 1506

    242 aProSerProAlaTyrSerSerSerSer..............LeuTyrA 254
   1507 ACCTGCCGACTTCCTCAGTCTGAGCAGTGAGACCAAGCCCAAGTTGATGA 1556

    254 laProAspAlaValAlaSerSerGlyProIleIleSerAspIleThrGlu 270
   1557 CACCTGACGCCTTCATGACACCTAGCGCCTCCTTGCAGCAGATCACT... 1603

    271 LeuAlaProAlaSerProMetAlaSerProGlyGlySerIleAspGluAr 287
   1604 .........GCCTCTCCCAGCAGCAGCAGCAGCGGTAGCAGCAGCAGCAG 1644

    287 gProLeuSerSerSerProLeuValArgVal.................L 298
   1645 CAGCAGTAGCAGCAGCTCCCTTACAGCTGTGTCTGCCATGAGCAGCACCT 1694

    298 ysGluGluProProSer.........ProProGln............Ser 307
   1695 CAGCTGTGGACCCCTCCTTGACCAGGCCACCTGAGGAGCTGACCTTGAGC 1744

    308 ProArgValGlu...GluAlaSerProGlyArgProSerSerValAspTh 323
   1745 CCCAAGCTGCAGCTGGATGGCAGCCTGACAATGAGCAGCAGTGGCAGCCT 1794

    323 rLeuLeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluP 340
   1795 TCAGGCAAGCCCGCGTGGCCTCCTGCCTGGCCTGCTC............C 1832

    340 roAlaProAla..... 343
```

```
    1833 CAGCCCCAGCTGACAAACTGACTCCCAAGGGGCCGGGCCAGGTGCCTACT 1882

     344 SerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyAr 360
         :::::::::||||||:::            :::::::::      |||
    1883 GCCACCTCTGCACTGTCCCTGGAGCTGCAGGAAGTGGAGCCCCTGGGGCT 1932

     360 gProProSerProProProThrSerThrProGluLysCysLeuSerValA 377
          |||   :::   |||      :::                 :::|||   |
    1933 ACCCCAAGCCTCCCCTAGCCGCACTCGTTCCCCTGATGTCATCTCCTCAG 1982

     377 laCysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSer 393
         ||      |||                              :::::::::
    1983 CTTCCACTGAC.............................CTGTCCCAG 2002

     394 AsnLeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAs 410
         ::::::   ::::::   :::          |||:::||||||
    2003 GACATCCCTGAGATTGCATCTGAGGCCCTGTCCCGTGGTTTT........ 2044

     410 pThrSerAlaLeuLeuAspLeuPheSerProSerValThrValProAspM 427
                                  :::   ::::::|||:::
    2045 ............................GGCTCCTCTGCACCAGAGG 2063

     427 etSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLeu 443
               ||||||      :::|||   ||||||      :::|||||||
    2064 GCCTTGAGCCAGACAGTATGGCTTCAGCCGCCTCGGCACTGCACCTGCTG 2113

     444 SerProGlnGluProProArgProProGluAlaGluAsnSerSerProAs 460
         ||||||:::      |||||||||                  :::|||::
    2114 TCCCCACGG......CCCCGGCCA.................GGGCCCGA 2139

     460 pSerGlyLysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuA 477
          :  |||  ||||||                              ||||
    2140 GCTCGGCCCCCAGCTCGGG.........................CTTG 2162

     477 spProGlySerValAspThrGlySerAsnAspLeuProValLeuPheGlu 493
         ||   |||      |||         ::::::   |||   |||:::|||
    2163 ATGGAGGCCCTGGGGATGGAGATCGGCATAATACCCCCTCCCTCCTGGAG 2212

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-764-233A-1

seq_documentation_block:
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homolog
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs t
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS ge
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  40190..46318
;      OTHER INFORMATION:  /product= "Module 5 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  46851..47891
```

```
;       OTHER INFORMATION:  /product= "SorM"
;       OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;       OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;       OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;       OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1


alignment_scores:
            Quality:  155.00              Length:    416
              Ratio:  0.799                Gaps:     18
 Percent Similarity:  46.635   Percent Identity: 26.442

alignment_block:
 US-09-304-121-2 x US-08-764-233A-1   ..

 Align seg 1/1  to: US-08-764-233A-1  from: 1  to: 49377

     139 LysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAs 155
         |||   |||   |||      |||   |||::::::
    8598 AAGCGGCTCCGCGACCACCTCCTCGCCAAAAGCGAG.............. 8633

     155 pSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluV 172
                 ||||||::                 :::|||
    8634 ..........CTCGCCCTCGCCGACGTGGCCTATTCGCTCGCGACCACGC 8673

     172 alAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLys 188
           |||   ::::::|||:::   |||         |||          ::::::
    8674 GCGCCCACTTCGAGCAGCGCGCCGCTCTCCTCGTCAAAGGCCGCGACGAG 8723

     189 LeuIleGlnPheLeuIleSerLeuValGln......SerAsnArgIleLe 203
         |||::::::   |||   ||||||:::|||      |||       :::||
    8724 CTCCTCTCCGCCCTCGATTCGCTGGCCCAAGGACATTCCGCCGCCGTGCT 8773

     203 u.........GlyValLysArgLysIleProLeuMetLeuAsnAspSerG 217
         |         |||:::      |||:::   ::::::::::::   :::|
    8774 CGGACGAAGCGGCGCCCCAGGAAAGCTCGCCGTCCTCTTCACGGGGCAAG 8823

     217 lySerAlaHis...SerMetProLysTyr.SerArgGlnPheSerLeuGl 232
         |||||   :::   :::|||   :::    |||      :::|||
    8824 GAAGCCAGCGGCCCACCATGGGCCGCGCCCTCTACGACGCTTTCC..... 8868

     232 uHisValHisGlySerGlyProTyrSerAlaProSerProAlaTyrSerS 249
                 :::||||||      |||:::||||||||||      |||:
    8869 .......CCGTCTTCCGGGACGCCCTCGACACCGTCGCCGCCCACCTCGA 8911

     249 erSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyProIleIle 265
         ::::::::      ||||||   |||::::::||||||   |||
    8912 CCGCGACCTCGACCGCCCCCTGCGCGACGTCCTCTTCGCTCCCG...... 8955

     265 SerAspIleThrGluLeuAlaProAlaSerProMetAlaSerProGlyGl 282
                       ||||||:::   |||   ||||||
    8956 ..............ACGGCTCCGAGCAGGCCGCGCGCCTCG........ 8982

     282 ySerIleAspGluArgProLeuSerSerSerProLeuValArgValLysG 299

    8983 .................................................A 8983

     299 luGluProProSerProProGlnSerProArgValGluGluAlaSerPro 315
           :::||||||||||||   :::   |||   :::   :::   ||||||
    8984 CCAAACCGCCTTCACCCAGCCGG...CCCTGTTTGCCCTCGAAGTCGCCC 9030

     316 GlyArgProSerSerValAspThrLeuLeuSerProThrAlaLeuIleAs 332
         :::        |||      ::::::   ||||||   :::      ||
    9031 TCTTTCAGCTTCTCCAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGA 9080

     332 pSerIleLeuArgGluSerGluPro......AlaProAlaSerValThrA 347
         |:::   |||   :::|||:::|||      :::||||||||||      :
    9081 CACTCCATTGGCGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCT 9130

     347 laLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProProSer 363
         ::   |||   |||                              |||   |||
    9131 CCAGGACGCCTGCA.........................CCCTCGTCG 9153

     364 ProProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAs 380
         ||||||      |||:::   ::::::      ::::::|||      ::
    9154 CCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTC 9203

     380 pLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn......L 395
         :   :::|||      :::   :::|||      |||:::      :
    9204 ACCCTCCGAGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGA 9253

     395 euAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThr 411
         ::|||:::|||   :::         |||   |||:::|||      :::
    9254 TGGACGAGCTAGCCTCGCCG...CCCTCAATGGGCCTCTCTCCACCGTCG 9300

     412 SerAlaLeuLeuAspLeu.........PheSerProSerValThrValPr 425
         |||       :::                |||||::          ||
    9301 TCGCTGGCGATGAAGACGCGGTGGTGGAGATCGCCCGCCAGGCCGAAGCC 9350

     425 oAspMetSerLeuProAspLeuAspSerSerLeu................. 436
         |:::   :::   |||          :::|||
    9351 CTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTTCCACTCTCC 9400

     437 .............AlaSerIleGlnGluLeuLeuSerProGlnGluPro 448
                       :::|||      :::      ||||||:::
    9401 GCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 9450

     449 ProArgProProGluAlaGluAsnSerSerProAspSerGlyLysGlnLe 465
         |||       |||:::|||:::   |||||||||   |||      :::
    9451 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGC 9500

     465 uValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValA 482
           :::   |||::::::|||                  ::::::
    9501 GCCACGGACCACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCG 9550

     482 spThr......GlySerAsnAspLeuProValLeuPhe...GluLeuGly 495
           |||      :::|||:::         |||   |||         |||
    9551 CCACACCGTCCGCTTCCTCGACGGCGTACGTGCCCTTCACGCCGAAGGGG 9600

     496 GluGlySerTyrPheSerGluGlyAspGlyPheAlaGluAspProThrIl 512
         ::::::|||:::   |||:::|||
    9601 CACGCGTCTTTCTCGAGCTCGGGC........................... 9624

     512 eSerLeuLeuThrGlySerGluProProLysAlaLysAspProThr 527
               ||||||  |||:::||||||      |||   |||:::
    9625 .......CTCACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCG 9663

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-258-261B-6

seq_documentation_block:
; Sequence 6, Application US/08258261B
; Patent No. 5639949
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,261B
;      FILING DATE:  08-JUN-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6


alignment_scores:
            Quality:  146.50              Length:    356
              Ratio:  0.899                Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-258-261B-6   ..

 Align seg 1/1  to: US-08-258-261B-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::      |||      |||:::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                        |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
           ||||||||||||        ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.........TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                     :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
         :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
                 ::::::   :::|||         ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |      ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
           :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                        ||
   10430 .....................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||         |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
               ||||||         ::::::      ||||||:::||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                    :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
             :::   ||||||:::|||:::         :::|||::::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::||||||||:::|||::::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
               :::|||:::|||:::|||                  ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
               |||         :::::::||::::::          :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
         ||                  |||          :::|||      :::
   10761 GC.................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                                 |||:::|||           :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
         :||||||   :::|||               ||||||:::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
         ||          |||||||||:::   :::   |||   :::      :::
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
         :::         |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
           |||   :::||||||
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-456-837-6

seq_documentation_block:
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
```

```
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/456,837
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-456-837-6


alignment_scores:
            Quality:  146.50              Length:    356
              Ratio:  0.899                Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-456-837-6   ..

 Align seg 1/1  to: US-08-456-837-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::       |||      |||::::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                            |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
             ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.........TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                    :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
          :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
                 ::::::   :::|||         ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |      ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
             :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                        ||
   10430 .....................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         |||||||||      |||         |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
             ||||||         ::::::      |||||||:::|||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                     :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
           :::   ||||||:::|||:::          :::|||::::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::||||||||||:::|||::::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
             :::|||:::|||:::|||                  ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
                |||          ::::::|||::::::          :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
         ||                  |||         :::|||      :::
   10761 GC.................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                                    |||:::|||           :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
         :||||||   :::|||           |||||||::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
         ||         |||||||||:::   :::   |||   :::      :::
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
         :::         |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
          |||   :::||||||
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-342-6

seq_documentation_block:
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
```

```
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6


alignment_scores:
            Quality:  146.50              Length:    356
              Ratio:  0.899                 Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-457-342-6   ..

 Align seg 1/1  to: US-08-457-342-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::      |||      |||:::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                        |||:::|||:::       :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
           ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.........TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                    :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
          :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
                 ::::::   :::|||          ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |      ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
           :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                        ||
   10430 .............................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||        |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
           ||||||        ::::::      ||||||:::||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                   :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
           :::   ||||||:::|||:::          :::|||::::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::|||||||||:::|||::::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
               :::|||:::|||:::|||                  ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
                 |||          ::::::|||::::::         :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
         ||                  |||         :::|||      :::
   10761 GC.................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                               |||:::|||            :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
         :||||||   :::|||             ||||||:::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
         ||        ||||||||||:::   :::   |||   :::      :::
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
         :::          |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
          |||   :::||||||
   10960 TTCCAGCGAGAGCGGT 10975
```

```
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-646A-6

seq_documentation_block:
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6


alignment_scores:
            Quality:  146.50              Length:    356
              Ratio:  0.899                Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-457-646A-6   ..

 Align seg 1/1  to: US-08-457-646A-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::      |||      |||::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165

         :                          |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
           ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.........TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                    :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
          :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
               ::::::   :::|||          ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |      ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
           :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                          ||
   10430 .....................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||         |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
           ||||||          ::::::      ||||||:::||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                   :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
           :::   ||||||:::|||:::          :::|||:::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::||||||||||:::|||::::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
               :::|||:::|||:::|||                  ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
               |||         ::::::|||::::::          :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
         ||                |||          :::|||      :::
   10761 GC.................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                              |||:::|||             :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
         :||||||   :::|||              ||||||:::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
         ||         ||||||||||:::   :::   |||   :::      :::
```

```
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
         :::          |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
         |||   :::|||||||
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-458-076A-6

seq_documentation_block:
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6


alignment_scores:
           Quality:  146.50            Length:    356
             Ratio:  0.899              Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157
```

```
alignment_block:
 US-09-304-121-2 x US-08-458-076A-6   ..

 Align seg 1/1  to: US-08-458-076A-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::       |||      |||:::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                          |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
         ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.........TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                    :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
          :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
                ::::::   :::|||          ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |     ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
              :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                        ||
   10430 .........................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||        |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
              ||||||         ::::::      ||||||:::||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                    :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
             :::   ||||||:::|||:::         :::|||:::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::||||||||:::|||:::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
              :::|||:::|||:::|||                  ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
                |||         ::::::|||::::::         :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
         ||                  |||         :::|||      :::
   10761 GC.................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789
```

```
     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                              |||:::|||             :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
         :||||||   :::|||               ||||||:::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
         ||         |||||||||:::   :::   |||   :::      :::
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
         :::         |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
          |||   :::||||||
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-764-233A-4

seq_documentation_block:
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:

;      CLONE:  p98/1
US-08-764-233A-4


alignment_scores:
            Quality:  146.50              Length:    356
              Ratio:  0.899                Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-764-233A-4   ..

 Align seg 1/1  to: US-08-764-233A-4  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::      |||      |||::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                          |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
           ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.........TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                     :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
         :::|||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
                 ::::::   :::|||          ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |     ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
             :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                          ||
   10430 .........................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||         |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
             ||||||          ::::::      ||||||:::||||||::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                     :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
           :::   ||||||:::|||:::          :::|||::::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::||||||||||:::|||::::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
             :::|||:::|||:::|||                  ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710
```

```
     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
                 |||         :::::::|||::::::         :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
          ||                  |||         :::|||      :::
   10761 GC..................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                              |||:::|||             :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
          :||||||   :::|||               ||||||:::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
          ||         |||||||||:::   :::   |||   :::       :::
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
          :::         |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
          |||   :::||||||
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-457-335A-6

seq_documentation_block:
; Sequence 6, Application US/08457335A
; Patent No. 5723759
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,335A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-335A-6

alignment_scores:
           Quality:  146.50              Length:    356
             Ratio:  0.899                 Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-457-335A-6   ..

 Align seg 1/1  to: US-08-457-335A-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::       |||       |||:::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                              |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
          ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.....TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                     :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
          :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
               ::::::   :::|||         ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |      ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
          :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                          ||
   10430 .........................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||         |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
          ||||||         ::::::      ||||||:::||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                 :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561

     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
```

```
           :::   ||||||:::|||:::          :::|||::::::   :
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaSerValThrAlaLeuThrAspAlaArgGly 353
         ::   :::|||||||||:::|||:::::::      :::   ||||||
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro.................ProSerPr 364
               :::|||:::|||:::|||                 ||||||
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
               |||          ::::::|||::::::          :::   |
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
         ||                   |||         :::|||      :::
   10761 GC.................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
                                 |||:::|||            :::::
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
         :||||||   :::|||           ||||||:::   :::   |
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
         ||         ||||||||||:::   :::   |||   :::      :::
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
         :::         |||      |||   |||   |||:::      :::
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
          |||   :::||||||
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-729-214-6

seq_documentation_block:
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6


alignment_scores:
            Quality:  146.50              Length:    356
              Ratio:  0.899                Gaps:     15
 Percent Similarity:  45.787   Percent Identity: 24.157

alignment_block:
 US-09-304-121-2 x US-08-729-214-6   ..

 Align seg 1/1  to: US-08-729-214-6  from: 1  to: 28958

     132 IleArgGlnAspSerValThrLysLeuLeuThrAspValGlnLeuMetLy 148
         :::|||:::|||:::   :::      |||      |||:::::::::::
   10081 CTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTCGACGTCGTACA 10130

     148 sGlyLysGlnGluCysMetAspSerLysLeuLeuAlaMetLysHisGluA 165
         :                           |||:::|||:::      :::
   10131 G.....................CCCGCCCTCTTTGCCGTCATGGTCTCCC 10159

     165 snGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLysHisAla... 180
           ||||||||||||         ||||||      :::   |||
   10160 TGGCCGCCCTCTGGCGC.....TCGCTCGGCGTCGAGCCCGCCGCC 10200

     181 ............GlnGlnGlnLysValValAsnLysLeuIleGlnPheLe 193
                     :::|||   :::::::::      ::::::
   10201 GTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTCGTCGCAGGCGC 10250

     193 uIleSerLeuValGlnSerAsnArgIleLeuGlyValLysArgLysIleP 210
          :::||||||   ::::::   ||||||   :::::::::||||||
   10251 TCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCGC 10300

     210 roLeuMetLeuAsnAspSerGly.........Ser.AlaHisSerMetPr 223
               ::::::   :::|||          ::: :::   |||   ||
   10301 TCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCC 10350

     223 oLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGlyProT 240
         |      ||||||      ||||||      :::   ||||||   |||
   10351 TCCGACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCACCGC 10400

     240 yrSerAlaProSerProAlaTyrSerSerSerSerLeuTyrAlaProAsp 256
           :::   :::|||:::      |||
   10401 CGCCGTCAACAGCCCCAGGGCTACCCTCG..................... 10429

     257 AlaValAlaSerSerGlyProIleIleSerAspIleThrGluLeuAlaPr 273
                                                          ||
   10430 .....................................TATCC 10434

     273 oAlaSerProMetAlaSerProGlyGlySerIleAspGluArgProLeuS 290
         ||||||||||      |||         |||   ::::::   |||
   10435 GGCGAGCCCGCCGCCGTCGACGCGCTGCTCGACGTCCTCACCGCCACCAA 10484

     290 erSerSerProLeuValArgValLysGluGluProProSerProProGln 306
           ||||||         ::::::      ||||||:::||||||:::
   10485 GGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCCGCCCAGA 10534

     307 SerProArgValGluGluAlaSerProGlyArgProSerSerValAspTh 323
                 :::   :::||||||
   10535 TGGACGCCGTCCAAGACGAGCTCGCCG..................... 10561
```

```
     323 rLeuLeuSerProThrAlaLeuIle.........AspSerIleLeuArgG 337
   10562 .CAGGTCTAGCCAACATCGCTCCTCGGACGTGCGAGCTCCCTCTTTATTC 10610

     337 luSerGluProAlaProAlaServalThrAlaLeuThrAspAlaArgGly 353
   10611 GACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTCGACGGCGCGTACT 10660

     354 HisThrAspThrGluGlyArgPro..................ProSerPr 364
   10661 GGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAGCGG 10710

     364 oProProThr.........SerThrProGluLysCysLeuSerValAlaC 378
   10711 CTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGT 10760

     378 ysLeuAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsn 394
   10761 GC..................TCACGCTCGCCCTCCGCGAGACCTGCG... 10789

     395 LeuAspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspTh 411
   10790 .......................AGCGCTCACCGCTCGATCCCGTCGTC 10815

     411 rSerAlaLeuLeuAsp.........LeuPheSerProSerValThrValP 425
   10816 GTCGGCTCCATTCGACGAGAAGAAGGCCACCTCGCCCGCCTGCTCCTCTC 10865

     425 roAsp...MetSerLeuProAspLeuAspSerSerLeuAlaSerIleGln 440
   10866 CTGGGCGGAGCTCTCTACCCGAGGCCTCGCGCTCGACTGGAAGGACTTCT 10915

     441 GluLeuLeuSerProGlnGluProProArgProProGluAlaGluAsnSe 457
   10916 TCG......CGCCCTACGCTCCCCGCAAGGTCTCCCTCCCCACCTACCCC 10959

     457 rSerProAspSerGly 462
   10960 TTCCAGCGAGAGCGGT 10975

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-864-038A-1

seq_documentation_block:
; Sequence 1, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
US-08-864-038A-1


alignment_scores:
            Quality:  145.00            Length:    331
              Ratio:  1.036               Gaps:     12
 Percent Similarity:  42.296   Percent Identity: 23.565

alignment_block:
 US-09-304-121-2 x US-08-864-038A-1/rev   ..

 Align seg 1/1  to reverse of: US-08-864-038A-1  from: 1  to: 2214

     145 GlnLeuMetLysGlyLysGlnGlu......Cys...........MetAs 155
     988 CAGCTGCAGCAGCGGCAGCAGCAGCACGTCTGCGGCCTCCTCTTCCTCGA 939

     155 pSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluV 172
     938 CCGCGTCTACCTCCTCGTCCACCGTAGAAACCAA......CACCACCGAG 895

     172 alAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLys 188
     894 TCCTCCACCTCCGGCAGCAGCGGCGGCAGCAGCAGCAGCGGCAGCAGCAG 845

     189 LeuIle..................GlnPheLeuIleSerLeuValGlnSe 199
     844 ATCCTCCATATCCTCCGAGGCCACCAAGACCTCCGAGGCCTCCAAGTCCG 795

     199 rAsnArgIleLeuGlyValLysArgLysIle..ProLeuMetLeuAsnAs 215
     794 CCACCGAGTCCTCCAAGTCCGCCACCGAGTCCTCCA.............. 759

     215 pSerGlySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuG 232
     758 ................AGTCCACCTGCGCCTCCTGCAGCGGCAGCAGCGG 725

     232 luHisValHisGlySerGlyProTyrSer......AlaProSerProAla 246
     724 CTGCGGCTGCAGCAGCAGCTCCTCCAAGTCTACCTGCGCCTCCTCCTGCA 675

     247 TyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyPr 263
     674 GCGGCAGCAGCGGCTGCGGCTGCGGCAGCAGCTCCACCAACTCCACCCCC 625

     263 oIleIleSerAspIleThrGluLeuAlaProAlaSerProMetAlaSerP 280
     624 GGCGGCAGCAGCGGCAGCTGCAGCAGCAGCAGCAGCACCTCCAGCTCCAC 575

     280 roGly...GlySerIleAspGluArgProLeuSerSerSerProLeuVal 295
     574 CAGCTCCACCGAGGGCAAGAGCTGCACCAAGATCATCGAAATCTAAATCA 525

     296 ArgValLysGluGluProProSerProProGlnSerProArgValGluGl 312
     524 AATAAATCGTCTCCACCGCCAAGTCCGCCAAGACCTCCAAGTCCGCCAAG 475

     312 uAlaSerProGlyArgProSerSerValAspThrLeuLeuSerProThrA 329
```

```
      :      |||      |||   :::                |||||||:::
    474 TCCACCTCCGAGACCTCCGCCCAATCCA.............AGTCCGAGT. 438

    329 laLeuIleAspSerIleLeuArgGluSerGluProAlaProAlaSerVal 345
                                    ||||||||||||   :::
    437 ...............................CCAGCACCTGCTCCTGCT 420

    346 ThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPr 362
          |||   :::   |||   :::   :::          :::   |||||
    419 CCTGCGCCTGCTCCAGCACCAGCACCAGCACCACCTCCGGCACCACCACC 370

    362 oSerProProProThrSerThrProGluLysCysLeuSerValAlaCysL 379
        |:::||||||||||:::   :::|||
    369 AGCACCACCTCCGGCGCCAGCTCCC........................... 345

    379 euAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeu 395

    345 ............................................... 345

    396 AspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThrSe 412

    345 ............................................... 345

    412 rAlaLeuLeuAspLeuPheSerProSerValThrValProAspMetSerL 429
                      :::   ||||||   :::          |||
    344 ..............ATTCCACCGTCATCATCCCAGTCCCATTCATCAT 311

    429 euProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLeuSerPro 445
          :::|||:::||||||   :::      |||         ||||||
    310 CGTCATCGCTGTCGTCATCATAGTCATATTCCCAA.........TCTCCA 270

    446 GlnGluProProArgProProGluAlaGluAsnSerSerPro 459
          ||||||||||||||:::|||      :::   |||
    269 CCGCCGCCGCCGCGTCCGCCGTCGGCTCCTCCAGCACCGCCA 228

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-864-038A-2

seq_documentation_block:
; Sequence 2, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3331
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
;    FEATURE:  mRNA
;      LOCATION:  from 1 to 3331
;      IDENTIFICATION METHOD:  E (by experiment)
US-08-864-038A-2

alignment_scores:
            Quality:  145.00            Length:    331
              Ratio:  1.036               Gaps:     12
 Percent Similarity:  42.296   Percent Identity: 23.565

alignment_block:
 US-09-304-121-2 x US-08-864-038A-2/rev   ..

 Align seg 1/1  to reverse of: US-08-864-038A-2  from: 1  to: 3331

     145 GlnLeuMetLysGlyLysGlnGlu......Cys.....MetAs 155
         ||||||:::::   :::|||:::      |||          :::||
    1037 CAGCTGCAGCAGCGGCAGCAGCAGCACGTCTGCGGCCTCCTCTTCCTCGA 988

     155 pSerLysLeuLeuAlaMetLysHisGluAsnGluAlaLeuTrpArgGluV 172
         |         |||   :::   :::::|||:::         ::::::|
     987 CCGCGTCTACCTCCTCGTCCACCGTAGAAACCAA......CACCACCGAG 944

     172 alAlaSerLeuArgGlnLysHisAlaGlnGlnGlnLysValValAsnLys 188
         ||      |||||||||:::::   |||||||||:::      ::::::
     943 TCCTCCACCTCCGGCAGCAGCGGCGGCAGCAGCAGCAGCGGCAGCAGCAG 894

     189 LeuIle.................GlnPheLeuIleSerLeuValGlnSe 199
         ::::::                 |||   |||   :::|||
     893 ATCCTCCATATCCTCCGAGGCCACCAAGACCTCCGAGGCCTCCAAGTCCG 844

     199 rAsnArgIleLeuGlyValLysArgLysIle..ProLeuMetLeuAsnAs 215
          :::|||:::|||   |||:::::::::::::  |||
     843 CCACCGAGTCCTCCAAGTCCGCCACCGAGTCCTCCA.............. 808

     215 pSerGlySerAlaHisSerMetProLysTyrSerArgGlnPheSerLeuG 232
                       |||   |||                 :::
     807 ................AGTCCACCTGCGCCTCCTGCAGCGGCAGCAGCGG 774

     232 luHisValHisGlySerGlyProTyrSer......AlaProSerProAla 246
           :::   ::::::::::|||   |||      ||||||   ||||||
     773 CTGCGGCTGCAGCAGCAGCTCCTCCAAGTCTACCTGCGCCTCCTCCTGCA 724

     247 TyrSerSerSerSerLeuTyrAlaProAspAlaValAlaSerSerGlyPr 263
           ::::::::::::      |||      |||      :::      ||
     723 GCGGCAGCAGCGGCTGCGGCTGCGGCAGCAGCTCCACCAACTCCACCCCC 674

     263 oIleIleSerAspIleThrGluLeuAlaProAlaSerProMetAlaSerP 280
         |     :::      :::       |||   |||:::|||   |||   |
     673 GGCGGCAGCAGCGGCAGCTGCAGCAGCAGCAGCAGCACCTCCAGCTCCAC 624

     280 roGly...GlySerIleAspGluArgProLeuSerSerSerProLeuVal 295
         ||:::                     |||   ||||||:::
     623 CAGCTCCACCGAGGGCAAGAGCTGCACCAAGATCATCGAAATCTAAATCA 574

     296 ArgValLysGluGluProProSerProProGlnSerProArgValGluGl 312
         :::   :::::   ||||||||||||||||:::   |||         ::
     573 AATAAATCGTCTCCACCGCCAAGTCCGCCAAGACCTCCAAGTCCGCCAAG 524

     312 uAlaSerProGlyArgProSerSerValAspThrLeuLeuSerProThrA 329
         :      |||      |||   :::             |||||||:::
     523 TCCACCTCCGAGACCTCCGCCCAATCCA.............AGTCCGAGT. 487
```

```
329 laLeuIleAspSerIleLeuArgGluSerGluProAlaProAlaSerVal 345
                                  ||||||||||||   :::
486 ..............................CCAGCACCTGCTCCTGCT 469

346 ThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPr 362
     |||   :::   |||   :::   :::         :::   |||||
468 CCTGCGCCTGCTCCAGCACCAGCACCAGCACCACCTCCGGCACCACCACC 419

362 oSerProProProThrSerThrProGluLysCysLeuSerValAlaCysL 379
    |:::||||||||||:::   :::|||
418 AGCACCACCTCCGGCGCCAGCTCCC..................... 394

379 euAspLysAsnGluLeuSerAspHisLeuAspAlaMetAspSerAsnLeu 395

394 ..................................................... 394

396 AspAsnLeuGlnThrMetLeuSerSerHisGlyPheSerValAspThrSe 412

394 ..................................................... 394

412 rAlaLeuLeuAspLeuPheSerProSerValThrValProAspMetSerL 429
                  :::   ||||||   :::            |||
393 ..............ATTCCACCGTCATCATCCCAGTCCCATTCATCAT 360

429 euProAspLeuAspSerSerLeuAlaSerIleGlnGluLeuLeuSerPro 445
       :::|||:::||||||   :::      |||         ||||||
359 CGTCATCGCTGTCGTCATCATAGTCATATTCCCAA.........TCTCCA 319

446 GlnGluProProArgProProGluAlaGluAsnSerSerPro 459
       ||||||||||||||:::|||      :::   |||
318 CCGCCGCCGCCGCGTCCGCCGTCGGCTCCTCCAGCACCGCCA 277
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2000, 00:11:14 ; Search time 45.51 Seconds
(without alignments)
5671.789 Million cell updates/sec

Title: US-09-304-121-1
Perfect score: 2156
Sequence: 1 cgggcccgttgcaagatggc.....aaaaaaaaaaaaaaaaaaaa 2156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2156	100.0	2156	2	US-08-178-477B-31	Sequence 31, Appl
	2	183.4	8.5	2781	2	US-08-178-477B-42	Sequence 42, Appl
	3	46.8	2.2	2338	2	US-08-425-069-1	Sequence 1, Appli
	4	46.8	2.2	2338	4	US-08-317-844B-1	Sequence 1, Appli
	5	46.4	2.2	2517	2	US-08-306-691B-18	Sequence 18, Appl
	6	46.4	2.2	2517	2	US-08-385-142-2	Sequence 2, Appli
	7	46.4	2.2	2517	3	US-08-481-814A-1	Sequence 1, Appli
c	8	46	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
	9	45.4	2.1	20235	2	US-07-642-734C-3	Sequence 3, Appli
	10	44.4	2.1	1035	1	US-07-601-094-30	Sequence 30, Appl
	11	44.4	2.1	1035	1	US-08-012-735-30	Sequence 30, Appl
	12	44.4	2.1	1914	1	US-07-601-094-1	Sequence 1, Appli
	13	44.4	2.1	1914	1	US-08-012-735-1	Sequence 1, Appli
	14	44.2	2.1	2403	2	US-08-471-033-30	Sequence 30, Appl
	15	44.2	2.1	2403	3	US-08-471-044-30	Sequence 30, Appl
	16	44.2	2.1	2403	3	US-08-463-483A-30	Sequence 30, Appl
	17	44.2	2.1	2403	3	US-08-471-046A-30	Sequence 30, Appl
	18	44.2	2.1	2403	3	US-08-470-566B-30	Sequence 30, Appl
	19	44.2	2.1	2403	3	US-08-838-219B-7	Sequence 7, Appli
	20	44.2	2.1	2403	4	US-08-469-334-30	Sequence 30, Appl
	21	43.8	2.0	2889	2	US-08-537-002A-4	Sequence 4, Appli
	22	43.8	2.0	3600	2	US-08-537-002A-5	Sequence 5, Appli
	23	43.8	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	24	43.6	2.0	3468	1	US-07-951-715A-2	Sequence 2, Appli
	25	43.6	2.0	3468	1	US-07-951-715A-4	Sequence 4, Appli
	26	43.6	2.0	3468	3	US-08-459-448A-2	Sequence 2, Appli
	27	43.6	2.0	3468	3	US-08-459-448A-4	Sequence 4, Appli
	28	43.6	2.0	6192	3	US-08-479-537A-1	Sequence 1, Appli
	29	43.6	2.0	6449	3	US-08-479-537A-4	Sequence 4, Appli
	30	43.2	2.0	1931	4	US-09-130-114-2	Sequence 2, Appli
c	31	43	2.0	54	2	US-08-178-477B-1	Sequence 1, Appli
c	32	43	2.0	2830	4	US-09-010-928B-1	Sequence 1, Appli
	33	42.8	2.0	2241	3	US-08-838-219B-20	Sequence 20, Appl
	34	42.8	2.0	30001	1	US-08-125-468-1	Sequence 1, Appli
	35	42.8	2.0	30001	3	US-08-474-933-1	Sequence 1, Appli
	36	42.6	2.0	2370	3	US-08-838-219B-19	Sequence 19, Appl
c	37	42.6	2.0	2824	4	US-09-010-928B-3	Sequence 3, Appli
	38	42.2	2.0	1957	1	US-08-295-060-3	Sequence 3, Appli
	39	42	1.9	1910	6	PCT-US92-05532-1	Sequence 1, Appli
	40	42	1.9	2261	1	US-08-272-882D-1	Sequence 1, Appli
	41	41.6	1.9	1995	2	US-08-425-069-3	Sequence 3, Appli
	42	41.6	1.9	1995	4	US-08-317-844B-3	Sequence 3, Appli
	43	41.6	1.9	2492	2	US-08-139-937-13	Sequence 13, Appl
	44	41.6	1.9	2492	6	PCT-US93-11310-13	Sequence 13, Appl
	45	41.2	1.9	1950	4	US-08-377-440A-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT   1
US-08-178-477B-31
; Sequence 31, Application US/08178477B
; Patent No. 5756343
;  GENERAL INFORMATION:
;    APPLICANT:  WU, CARL; CLOS, JOACHIM;
;    APPLICANT:  WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
;    TITLE OF INVENTION:  CELL STRESS
;    TITLE OF INVENTION:  TRANSCRIPTIONAL FACTORS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/178,477B
;      FILING DATE:  07-JAN-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/617,910
;      FILING DATE:  26-NOV-1990
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4103US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2156
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  (DNA) genomic
US-08-178-477B-31


  Query Match            100.0%;  Score 2156;  DB 2;  Length 2156;
  Best Local Similarity  100.0%;  Pred. No. 0;
```

Matches 2156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 cgggcccgttgcaagatggcggcggccatgctgggccccggggctgtgtgtgcgcagcgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGGCCCGTTGCAAGATGGCGGCGGCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGG 60

Qy     61 gcggcggcgcggcccggaaggctggcgcggcgacggcgttagcccggccctcggcccctc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGGCGGCGCGGCCCGGAAGGCTGGCGCGGCGACGGCGTTAGCCCGGCCCTCGGCCCCTC 120

Qy    121 tttgcggccgctccctccgcctattccctccttgctcgagatggatctgcccgtgggccc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGCGGCCGCTCCCTCCGCCTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCC 180

Qy    181 cggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctcgtgag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGGCGCGGCGGGGCCCAGCAACGTCCCGGCCTTCCTGACCAAGCTGTGGACCCTCGTGAG 240

Qy    241 cgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccacgtgtt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGCGGGAACAGCTTCCACGTGTT 300

Qy    301 cgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaacatggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACAACAACATGGC 360

Qy    361 cagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgagcaggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGG 420

Qy    421 cggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctgcgtgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGG 480

Qy    481 ccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctgaagag 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCAGGAGCAGCTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAG 540

Qy    541 tgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcagctgat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGAAGACATAAAGATCCGCCAGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGAT 600

Qy    601 gaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaatgaggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAAGGGGAAGCAGGAGTGCATGGACTCCAAGCTCCTGGCCATGAAGCATGAGAATGAGGC 660

Qy    661 tctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCCCAGCAACAGAAAGTCGTCAA 720

Qy    721 caagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtgaagag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGGGGGTGAAGAG 780

Qy    781 aaagatccccctgatgctgaacgacagtggctcagcacattccatgcccaagtatagccg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCCG 840

Qy    841 gcagttctccctggagcacgtccacggctcgggcccctactcggccccctccccagccta 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTA 900

Qy    901 cagcagctccagcctctacgcccctgatgctgtggccagctctggacccatcatctccga 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAGCAGCTCCAGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGA 960

Qy    961 catcaccgagctggctcctgccagccccatggcctcccccggcgggagcatagacgagag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CATCACCGAGCTGGCTCCTGCCAGCCCCATGGCCTCCCCCGGCGGGAGCATAGACGAGAG 1020

Qy   1021 gcccctatccagcagccccctggtgcgtgtcaaggaggagcccccagcccgcctcagag 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCCCCTATCCAGCAGCCCCCTGGTGCGTGTCAAGGAGGAGCCCCCCAGCCCGCCTCAGAG 1080

Qy   1081 cccccgggtagaggaggcgagtcccgggcgcccatcttccgtggacaccctcttgtcccc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCCCCGGGTAGAGGAGGCGAGTCCCGGGCGCCCATCTTCCGTGGACACCCTCTTGTCCCC 1140

Qy   1141 gaccgccctcattgactccatcctgcgggagagtgaacctgcccccgcctccgtcacagc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GACCGCCCTCATTGACTCCATCCTGCGGGAGAGTGAACCTGCCCCCGCCTCCGTCACAGC 1200

Qy   1201 cctcacggacgccaggggccacacggacaccgagggccggcctccctccccccccgcccac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTCACGGACGCCAGGGGCCACACGGACACCGAGGGCCGGCCTCCCTCCCCCCCCGCCCAC 1260

Qy   1261 ctccacccctgaaaagtgcctcagcgtagcctgcctggacaagaatgagctcagtgacca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTCCACCCCTGAAAAGTGCCTCAGCGTAGCCTGCCTGGACAAGAATGAGCTCAGTGACCA 1320

Qy   1321 cttggatgctatggactccaacctggataacctgcagaccatgctgagcagccacggctt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTTGGATGCTATGGACTCCAACCTGGATAACCTGCAGACCATGCTGAGCAGCCACGGCTT 1380

Qy   1381 cagcgtggacaccagtgccctgctggacctgttcagcccctcggtgaccgtgcccgacat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACAT 1440

Qy   1441 gagcctgcctgaccttgacagcagcctggccagtatccaagagctcctgtctccccagga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GAGCCTGCCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGA 1500

Qy   1501 gccccccaggcctcccgaggcagagaacagcagcccggattcagggaagcagctggtgca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCA 1560

Qy   1561 ctacacagcgcagccgctgttcctgctggaccccggctccgtggacaccgggagcaacga 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGA 1620

Qy   1621 cctgccggtgctgtttgagctgggagagggctcctacttctccgaaggggacggcttcgc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGACGGCTTCGC 1680

Qy   1681 cgaggaccccaccatctccctgctgacaggctcggagcctcccaaagccaaggaccccac 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCAC 1740

Qy   1741 tgtctcctagaggccccggaggagctgggccagccgcccaccccccaccccccagtgcaggg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGTCTCCTAGAGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGG 1800

Qy   1801 ctggtcttggggaggcagggcagcctcgcggtcttgggcactggtgggtcggccgccata 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CTGGTCTTGGGGAGGCAGGGCAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATA 1860

Qy   1861 gccccagtaggacaaacgggctcgggtctgggcagcacctctggtcaggagggtcaccct 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCCCCAGTAGGACAAACGGGCTCGGGTCTGGGCAGCACCTCTGGTCAGGAGGGTCACCCT 1920

Qy   1921 ggcctgccagtctgccttcccccaaccccgtgtcctgtggtttggttggggcttcacagc 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGGTTTGGTTGGGGCTTCACAGC 1980

Qy   1981 cacacctggactgaccctgcaggttgttcatagtcagaattgtattttggatttttacac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGGATTTTTACAC 2040

Qy   2041 aactgtcccgttccccgctccacagagatacacagatatatacacacagtggatggacgg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACGG 2100

Qy   2101 acaagacaggcagagatctataaacagacaggctctaaaaaaaaaaaaaaaaaaaaa 2156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAAA 2156
```

```
RESULT   2
US-08-178-477B-42
; Sequence 42, Application US/08178477B
; Patent No. 5756343
;  GENERAL INFORMATION:
;    APPLICANT:  WU, CARL; CLOS, JOACHIM;
;    APPLICANT:  WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
;    TITLE OF INVENTION:  CELL STRESS
;    TITLE OF INVENTION:  TRANSCRIPTIONAL FACTORS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/178,477B
;      FILING DATE:  07-JAN-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/617,910
;      FILING DATE:  26-NOV-1990
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4103US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2781
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-178-477B-42


  Query Match             8.5%;  Score 183.4;  DB 2;  Length 2781;
  Best Local Similarity   57.8%;  Pred. No. 1.8e-30;
  Matches  366;  Conservative    0;  Mismatches  261;  Indels    6;  Gaps    2;

Qy    176 ggccccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgtggaccctc 235
          ||   || ||||||       |||   || ||||||||  || ||||  |||||  |||
Db    337 GGAGACGCCGCGGCCATCGGAAGCGGGGTGCCGGCCTTTTTGGCCAAATTGTGGCGCCTG 396

Qy    236 gtgagcgacccggacaccgacgcgctcatctgctggagcccgagcgggaacagcttccac 295
          |||  |||  | || ||| |     | || |||||||| |  |   ||  | || |||
Db    397 GTGGACGATGCCGATACCAATCGCTTGATTTGCTGGACCAAGGATGGCCAAAGTTTCGTT 456

Qy    296 gtgttcgaccagggccagtttgccaaggaggtgctgcccaagtacttcaagcacaacaac 355
          |     | || |  || ||||||||||  |  ||||     ||| ||||||||||||
Db    457 ATTCAAAATCAAGCGCAATTTGCCAAGGAACTATTGCCACTAAACTACAAGCACAACAAC 516

Qy    356 atggccagcttcgtgcggcagctcaacatgtatggcttccggaaagtggtccacatcgag 415
          |||||||| ||| |  ||||  | || |||||||| ||||  ||  |   |   || ||
Db    517 ATGGCCAGTTTCATAAGGCAATTGAATATGTATGGATTCCACAAGATCACCTCTATTGAC 576

Qy    416 cagggcggcctggtcaagccagagagagacgacacggagttccagcacccatgcttcctg 475
          | ||||| |   |       ||  | ||||| |  || ||   ||||||||  ||   |
Db    577 AATGGCGGAC---TACGTTTTGATCGCGACGAGATTGAATTTTCGCACCCATTTTTTAAG 633

Qy    476 cgtggccaggagcagctccttgagaacatcaagaggaaagtgaccagtgtgtccaccctg 535
          ||   |  |      || |||||  | ||||| ||||||  |  | |        |
Db    634 CGCAACTCGCCTTTTCTACTTGACCAAATCAAAAGGAAAATATCGAACAACAAAAATGGT 693

Qy    536 aagagtgaagacataaagatccgccaggacagcgtcaccaagctgctgacggacgtgcag 595
          |     |||   |   ||  |   ||| ||   |  | ||| | || || || ||| |
Db    694 GACGACAAAGGTGTCCTGAAGCCGGAGGCCA---TGTCGAAGATTCTCACCGATGTGAAA 750

Qy    596 ctgatgaaggggaagcaggagtgcatggactccaagctcctggccatgaagcatgagaat 655
          | |||  |||    |||||     |||| ||     ||   |||||||| || |||||
Db    751 GTCATGCGGGGTCGTCAGGACAATCTGGATTCGCGCTTCTCCGCCATGAAACAGGAGAAC 810

Qy    656 gaggctctgtggcgggaggtggccagccttcggcagaagcatgcccagcaacagaaagtc 715
          || |  |||||||| ||| | |||||||| || || ||||| ||  |||| ||  || |
Db    811 GAAGTGCTGTGGCGCGAGATAGCCAGCCTGCGCCAAAAGCACGCTAAGCAGCAACAAATA 870

Qy    716 gtcaacaagctcattcagttcctgatctcactggtgcagtcaaaccggatcctgggggtg 775
          |||||||| || || |||||||| ||  |  | |||||  |    || | | ||   |
Db    871 GTCAACAAACTGATCCAGTTCCTCATTACCATTGTGCAACCGTCGCGCAACATGTCTGGC 930

Qy    776 aagagaaagatccccctgatgctgaacgacagt 808
          || |        | | || ||| | ||| |
Db    931 GTGAAACGCCATGTGCAGCTGATGATCAACAAT 963


RESULT   3
US-08-425-069-1
; Sequence 1, Application US/08425069
; Patent No. 5728810
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5728810th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,069
;      FILING DATE:  19-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2338 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephilia clavipes
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  1..2154
;      OTHER INFORMATION:  /product= "Nephila clavipes
;      OTHER INFORMATION:  dragline silk protein"
;    PUBLICATION INFORMATION:
;      AUTHORS:  Xu, Ming
;      AUTHORS:  Lewis, Randolph V.
;      TITLE:  Structure of a protein superfiber:  Spider
;      TITLE:  drafline silk
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  87
;      PAGES:  7120-7124
;      DATE:  Sept.-1990
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 2338
US-08-425-069-1


  Query Match             2.2%;  Score 46.8;  DB 2;  Length 2338;
  Best Local Similarity   45.7%;  Pred. No. 0.14;
  Matches  202;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    1;

Qy    317 gccaaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcag 376
          | ||||||||    |    |  |  |   | ||||  |  |||         ||| ||||
Db    680 GACAAGGAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCATCCGCTGCAGCAG 739

Qy    377 ctcaacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagcca 436
          ||     ||   | |     | |  |||    | | ||      |  |||| | ||
Db    740 CTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCTGGACGAGGTG 799

Qy    437 gagagagacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctcctt 496
          ||||  |   |  || ||   ||||| |          |||  | | ||||| |     |
Db    800 GAGAAGGTGCAGGCGCAGCCGCAGCAGCAGCCGGAGGTGCTGGACAAGGAGGATACGGTG 859

Qy    497 gagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagaca---taaag 553
          |      |  | | |      |  ||  | | | |    | |    |||| ||   |   |
Db    860 GTCTTGGTGGACAAGGTGCCGGACAAGGAGGCTATGGAGGACTTGGAAGCCAAGGTGCTG 919

Qy    554 atccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcag 613
            |   ||||    ||   |||| ||| |  | |   ||||||  |  |  |    ||
Db    920 GACGAGGAGGATTAGGTGGACAAGGTGCAGGTGCAGCAGCAGCTGGAGGTGCCGGGCAAG 979

Qy    614 gagtgcatggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggag 673
          |||  |  ||    |||| | |||| || | |||  ||| |   |   | |  || || |
Db    980 GAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCAGCCGCTGCAGCAGCTGGTG 1039

Qy    674 gtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcag 733
          ||| |   |     | | | | | | |  | ||   || ||     || || |   | ||
Db   1040 GTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCAGGACGAGGTGGATTAG 1099

Qy    734 ttcctgatctcactggtgcagt 755
          |    |     | ||||||||
Db   1100 GTGGACAAGGGGCAGGTGCAGT 1121


RESULT   4
US-08-317-844B-1
; Sequence 1, Application US/08317844B
; Patent No. 5989894
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5989894th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,844B
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-105P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 241-1300
;      TELEFAX:  (703) 241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2338 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephilia clavipes
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2154
;      OTHER INFORMATION:  /product= "Nephila clavipes
;      OTHER INFORMATION:  dragline silk protein"
;    PUBLICATION INFORMATION:
;      AUTHORS:  Xu, Ming
;      AUTHORS:  Lewis, Randolph V.
;      TITLE:  Structure of a protein superfiber:  Spider
;      TITLE:  drafline silk
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  87
;      PAGES:  7120-7124
;      DATE:  Sept.-1990
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 2338
US-08-317-844B-1


  Query Match             2.2%;  Score 46.8;  DB 4;  Length 2338;
  Best Local Similarity   45.7%;  Pred. No. 0.14;
  Matches  202;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    1;

Qy    317 gccaaggaggtgctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcag 376
          | ||||||||    |    |  |  |   | ||||  |  |||         ||| ||||
Db    680 GACAAGGAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCATCCGCTGCAGCAG 739

Qy    377 ctcaacatgtatggcttccggaaagtggtccacatcgagcagggcggcctggtcaagcca 436
          ||     ||   | |     | |  |||    | | ||      |  |||| | ||
Db    740 CTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCTGGACGAGGTG 799

Qy    437 gagagagacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctcctt 496
          ||||  |   |  || ||   ||||| |          |||  | | ||||| |     |
Db    800 GAGAAGGTGCAGGCGCAGCCGCAGCAGCAGCCGGAGGTGCTGGACAAGGAGGATACGGTG 859

Qy    497 gagaacatcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagaca---taaag 553
          |      |  | | |      |  ||  | | | |    | |    |||| ||   |   |
Db    860 GTCTTGGTGGACAAGGTGCCGGACAAGGAGGCTATGGAGGACTTGGAAGCCAAGGTGCTG 919

Qy    554 atccgccaggacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcag 613
            |   ||||    ||   |||| ||| |  | |   ||||||  |  |  |    ||
Db    920 GACGAGGAGGATTAGGTGGACAAGGTGCAGGTGCAGCAGCAGCTGGAGGTGCCGGGCAAG 979

Qy    614 gagtgcatggactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggag 673
          |||  |  ||    |||| | |||| || | |||  ||| |   |   | |  || || |
Db    980 GAGGACTAGGTGGACAAGGTGCTGGACAAGGAGCTGGAGCAGCCGCTGCAGCAGCTGGTG 1039

Qy    674 gtggccagccttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcag 733
```

```
Db   1040 GTGCCGGACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCAGGACGAGGTGGATTAG 1099

Qy    734 ttcctgatctcactggtgcagt 755
Db   1100 GTGGACAAGGGGCAGGTGCAGT 1121
```

RESULT 5
US-08-306-691B-18
; Sequence 18, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-306-691B-18

Query Match 2.2%; Score 46.4; DB 2; Length 2517;
Best Local Similarity 49.2%; Pred. No. 0.17;
Matches 122; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy     38 ccggggctgtgtgtgcgcagcgggcggcggcgcggcccggaaggctggcgcggcgacggc 97
Db      7 CCGTGGCCGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGCGGGATCGAGCCCTCGCC 66

Qy     98 gttagcccggccctcggcccctctttgcggccgctccctccgcctattccctccttgctc 157
Db     67 GAGGCCTGCCGCCATGGGCCCGCGCCGCCGCCGCCGCCTGTCACCCGGGCCGCGCGGGCC 126

Qy    158 gagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctg 217
Db    127 GTGAGCGTCATGGCCTTGGCCGGGGCCCCTGCGGGCGGCCCATGCGCGCCGGCGCTGGAG 186

Qy    218 accaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccg 277
Db    187 GCCCTGCTCGGGGCCGGCGCGCTGCGGCTGCTCGACTCCTCGCAGATCGTCATCATCTCC 246

Qy    278 agcgggaa 285
Db    247 GCCGCGCA 254
```

RESULT 6
US-08-385-142-2
; Sequence 2, Application US/08385142
; Patent No. 5811299
; GENERAL INFORMATION:
; APPLICANT: Renner, Wolfgang A.
; APPLICANT: Eppenberger, Hans M.
; APPLICANT: Bailey, James E.
; TITLE OF INVENTION: Methods for the Activation of Proliferation
; TITLE OF INVENTION: of Animal Cells
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,142
; FILING DATE: 07-FEB-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Jeffrey S.
; REGISTRATION NUMBER: 31,879
; REFERENCE/DOCKET NUMBER: 27656/32470
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-385-142-2

Query Match 2.2%; Score 46.4; DB 2; Length 2517;
Best Local Similarity 49.2%; Pred. No. 0.17;
Matches 122; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy     38 ccggggctgtgtgtgcgcagcgggcggcggcgcggcccggaaggctggcgcggcgacggc 97
Db      7 CCGTGGCCGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGCGGGATCGAGCCCTCGCC 66

Qy     98 gttagcccggccctcggcccctctttgcggccgctccctccgcctattccctccttgctc 157
Db     67 GAGGCCTGCCGCCATGGGCCCGCGCCGCCGCCGCCGCCTGTCACCCGGGCCGCGCGGGCC 126

Qy    158 gagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctg 217
Db    127 GTGAGCGTCATGGCCTTGGCCGGGGCCCCTGCGGGCGGCCCATGCGCGCCGGCGCTGGAG 186

Qy    218 accaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccg 277
Db    187 GCCCTGCTCGGGGCCGGCGCGCTGCGGCTGCTCGACTCCTCGCAGATCGTCATCATCTCC 246

Qy    278 agcgggaa 285
```

```
Db    247 GCCGCGCA 254


RESULT   7
US-08-481-814A-1
; Sequence 1, Application US/08481814A
; Patent No. 5869040
;  GENERAL INFORMATION:
;    APPLICANT:  Hsu, Yen-Ming
;    TITLE OF INVENTION:  GENE THERAPY METHODS AND COMPOSITIONS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Biogen, Inc.
;      STREET:  14 Cambridge Center
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  02142
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,814A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kaplan, Warren A
;      REFERENCE/DOCKET NUMBER:  A001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-679-2000
;      TELEFAX:  617-679-2838
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2517 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      DEVELOPMENTAL STAGE:  Fetus
;      TISSUE TYPE:  Brain
;      CELL TYPE:  B-cell precursor
;      CELL LINE:  Nalm 6
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  136..1446
;      OTHER INFORMATION:  /function= "pRB binding protein"
;      OTHER INFORMATION:  /product= "E2F-1"
US-08-481-814A-1


  Query Match             2.2%;  Score 46.4;  DB 3;  Length 2517;
  Best Local Similarity   49.2%;  Pred. No. 0.17;
  Matches  122;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy      38 ccggggctgtgtgtgcgcagcgggcggcggcgcggcccggaaggctggcgcggcgacggc 97
           ||| ||| | |  | |||| ||||||||  ||  |||| |    || | |  || |
Db       7 CCGTGGCCGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGCGGGATCGAGCCCTCGCC 66

Qy      98 gttagcccggccctcggcccctctttgcggccgctccctccgcctattccctccttgctc 157
           |    |     ||  || ||| |   || ||||| |||    |     || |   |  |
Db      67 GAGGCCTGCCGCCATGGGCCCGCGCCGCCGCCGCCGCCTGTCACCCGGGCCGCGCGGGCC 126

Qy     158 gagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctg 217
           | ||  |   || || ||| |   | | | |||||   |     ||  |||||  |   |
Db     127 GTGAGCGTCATGGCCTTGGCCGGGGCCCCTGCGGGCGGCCCATGCGCGCCGGCGCTGGAG 186

Qy     218 accaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccg 277
           ||  |||  || ||  || |      | |  |  |  | |||  |||  |   | | |
Db     187 GCCCTGCTCGGGGCCGGCGCGCTGCGGCTGCTCGACTCCTCGCAGATCGTCATCATCTCC 246

Qy     278 agcgggaa 285
           || | |
Db     247 GCCGCGCA 254


RESULT   8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14


  Query Match             2.1%;  Score 46;  DB 1;  Length 7218;
  Best Local Similarity   6.6%;  Pred. No. 0.26;
  Matches   28;  Conservative  214;  Mismatches  184;  Indels    0;  Gaps    0;

Qy     329 ctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtat 388
           ||   |||||| || |||    | || |           : : :: ::   :: : : :
Db    1471 CTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRR 1412

Qy     389 ggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgac 448
           ::     :::::: ::   : :  ::: :::: ::   ::  :::  :::::::::: ::
Db    1411 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352

Qy     449 acggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaag 508
           : ::::    :: :   : :      : : ::  ::::: ::      ::::: :  :::
Db    1351 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1292
```

```
Qy     509 aggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagc 568
           ::::::: ::  :: : :   :    :::::: ::::: : :::::   :  :::: ::
Db    1291 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1232

Qy     569 gtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatggactcc 628
           :  :  :::  :  :: ::: : : ::  :: :::::::::: ::::: : : :::
Db    1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172

Qy     629 aagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagccttcgg 688
           :::     ::  : :::: : ::::: ::::    : :: :::::: ::  ::     ::
Db    1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112

Qy     689 cagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatctcactg 748
           ::::: : :   :: :: ::::::  :  :: :::   :   :: ||   | ||| ||
Db    1111 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTC 1052

Qy     749 gtgcag 754
           |  | |
Db    1051 GACCTG 1046


RESULT   9
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7165..9216
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9433..9984
;      OTHER INFORMATION:  /function= "approximate span
;      OTHER INFORMATION:  beta-ketoreductase of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10225..10483
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 4"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10723..20235
;      OTHER INFORMATION:  /codon_start= 10723
;      OTHER INFORMATION:  /function= "gene =eryA"
;      OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;      OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10723..15165
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10831..12174
```

```
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  12379..13350
;      OTHER INFORMATION:  /function= "approximatr span of
;      OTHER INFORMATION:  acyltransferase domain of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14062..14610
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14857..15114
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15166..20235
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15172..16569
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  16768..17721
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  18379..18921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19149..19398
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19492..20235
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3


  Query Match             2.1%;  Score 45.4;  DB 2;  Length 20235;
  Best Local Similarity   61.3%;  Pred. No. 0.45;
  Matches   73;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy     168 tgcccgtgggccccggcgcggcggggcccagcaacgtcccggccttcctgaccaagctgt 227
           |||  | || |||  |||  | || ||    |    ||| |||   |  || |||| ||
Db   13955 TGCTGGCGGACCCGCGCGGCGAGGAGCAGGTCGCGATCCGGGCGGACGGGATCAAGGTGG 14014

Qy     228 ggaccctcgtgagcgacccggacaccgacgcgctcatctgctggagcccgagcgggaac 286
           ||  ||||||  || ||| | |  || ||| | || | ||||||||||  |||| | |
Db   14015 CGAGGCTCGTGCCCGCCCCCGCCCGCGCCGCACGCACCCGCTGGAGCCCTCGCGGCACC 14073


RESULT  10
US-07-601-094-30
; Sequence 30, Application US/07601094
; Patent No. 5215892
;  GENERAL INFORMATION:
;    APPLICANT:  Kishimoto, Tadamitsu
;    APPLICANT:  Hirano, Toshio
;    APPLICANT:  Akira, Shizuo
;    APPLICANT:  Isshiki, Hiroshi
;    APPLICANT:  Tanabe, Osamu
;    APPLICANT:  Kinoshita, Shigemi
;    APPLICANT:  Shimamoto, Takuya
;    TITLE OF INVENTION:  C/EBP2 Gene and Recombinant
;    TITLE OF INVENTION:  C/EBP2
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak &
;      ADDRESSEE:  Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/601,094
;      FILING DATE:  19901022
;      CLASSIFICATION:  435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1035 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1035
;      OTHER INFORMATION:
US-07-601-094-30


  Query Match             2.1%;  Score 44.4;  DB 1;  Length 1035;
  Best Local Similarity   46.7%;  Pred. No. 0.36;
  Matches  141;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     443 gacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaac 502
           |||| || ||    || |  | | |||| |  | | | |  |  |  || |    ||  |
Db     718 GACGCCAAGGCCCCCCCGACCGCCTGCTACGCGGGGGCCGGGCCGGCGCCCTCGCAGGTC 777

Qy     503 atcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccag 562
           |  |  | |     ||  |  |||  ||  |  |     ||  |||  |     ||| ||
Db     778 AAGAGCAAGGCCAAGAAGACCGTGGACAAGCACAGCGACGAGTACAAGATCCGGCGCGAG 837

Qy     563 gacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatg 622
             || |  || |    |||  | |  |    |   |   |||  |    |   |  | |
Db     838 CGCAACAACATCGCCGTGCGCAAGAGCCGCGACAAGGCCAAGATGCGCAACCTGGAGACG 897

Qy     623 gactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagc 682
           |   |||| ||||||   | | |   ||||| |||   |||  |  | ||||||
Db     898 CAGCACAAGGTCCTGGAGCTCACGGCCGAGAACGAGCGGCTGCAGAAGAAGGTGGAGCAG 957

Qy     683 cttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatc 742
           ||   ||   |||    |   |  | |||  |  ||||  ||||     ||  |||| ||
Db     958 CTGTCGCGCGAGCTCAGCACCCTGCGGAACTTGTTCAAGCAGCTGCCCGAGCCCCTGCTC 1017

Qy     743 tc 744
           |
Db    1018 GC 1019


RESULT  11
US-08-012-735-30
; Sequence 30, Application US/08012735
; Patent No. 5360894
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Kishimoto, Tadamitsu
;    APPLICANT:  Hirano, Toshio
;    APPLICANT:  Akira, Shizuo
;    APPLICANT:  Isshiki, Hiroshi
;    APPLICANT:  Tanabe, Osamu
;    APPLICANT:  Kinoshita, Shigemi
;    APPLICANT:  Shimamoto, Takuya
;    TITLE OF INVENTION:  C/EBP2 Gene and Recombinant
;    TITLE OF INVENTION:  C/EBP2
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak &
;      ADDRESSEE:  Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/012,735
;      FILING DATE:  19930203
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/601,094
;      FILING DATE:  22 OCT 1990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1035 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1035
;      OTHER INFORMATION:
US-08-012-735-30


  Query Match             2.1%;  Score 44.4;  DB 1;  Length 1035;
  Best Local Similarity   46.7%;  Pred. No. 0.36;
  Matches  141;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     443 gacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaac 502
           |||| || ||    || |  | | |||| |  | | | |  |  |  || |    ||  |
Db     718 GACGCCAAGGCCCCCCCGACCGCCTGCTACGCGGGGGCCGGGCCGGCGCCCTCGCAGGTC 777

Qy     503 atcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccag 562
           |  |  | |     ||  |  |||  ||  |  |     ||  |||  |     ||| ||
Db     778 AAGAGCAAGGCCAAGAAGACCGTGGACAAGCACAGCGACGAGTACAAGATCCGGCGCGAG 837

Qy     563 gacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatg 622
             || |  || |    |||  | |  |     |   |||  |     |   |  |  | |
Db     838 CGCAACAACATCGCCGTGCGCAAGAGCCGCGACAAGGCCAAGATGCGCAACCTGGAGACG 897

Qy     623 gactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagc 682
             |   |||| ||||||    | | |   ||||| |||   |||  |  | ||||||
Db     898 CAGCACAAGGTCCTGGAGCTCACGGCCGAGAACGAGCGGCTGCAGAAGAAGGTGGAGCAG 957

Qy     683 cttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatc 742
           ||  ||  |||    |  |  | |||  |  ||||  ||||     ||  |||| ||
Db     958 CTGTCGCGCGAGCTCAGCACCCTGCGGAACTTGTTCAAGCAGCTGCCCGAGCCCCTGCTC 1017

Qy     743 tc 744
           |
Db    1018 GC 1019


RESULT  12
US-07-601-094-1
; Sequence 1, Application US/07601094
; Patent No. 5215892
;  GENERAL INFORMATION:
;    APPLICANT:  Kishimoto, Tadamitsu
;    APPLICANT:  Hirano, Toshio
;    APPLICANT:  Akira, Shizuo
;    APPLICANT:  Isshiki, Hiroshi
;    APPLICANT:  Tanabe, Osamu
;    APPLICANT:  Kinoshita, Shigemi
;    APPLICANT:  Shimamoto, Takuya
;    TITLE OF INVENTION:  C/EBP2 Gene and Recombinant
;    TITLE OF INVENTION:  C/EBP2
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak &
;      ADDRESSEE:  Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:.
;      APPLICATION NUMBER:  US/07/601,094
;      FILING DATE:  19901022
;      CLASSIFICATION:  435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1914 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  281..1316
;      OTHER INFORMATION:
US-07-601-094-1


  Query Match             2.1%;  Score 44.4;  DB 1;  Length 1914;
  Best Local Similarity   46.7%;  Pred. No. 0.42;
  Matches  141;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     443 gacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaac 502
           |||| || ||    || |  | | |||| |  | | | |  |  |  || |    ||  |
Db     998 GACGCCAAGGCCCCCCCGACCGCCTGCTACGCGGGGGCCGGGCCGGCGCCCTCGCAGGTC 1057

Qy     503 atcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccag 562
           |  |  | |     ||  |  |||  ||  |  |     ||  |||  |     ||| ||
Db    1058 AAGAGCAAGGCCAAGAAGACCGTGGACAAGCACAGCGACGAGTACAAGATCCGGCGCGAG 1117

Qy     563 gacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatg 622
             || |  || |    |||  | |  |     |   |||  |     |   |  |  | |
Db    1118 CGCAACAACATCGCCGTGCGCAAGAGCCGCGACAAGGCCAAGATGCGCAACCTGGAGACG 1177

Qy     623 gactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagc 682
             |   |||| ||||||    | | |   ||||| |||   |||  |  | ||||||
```

```
Db   1178 CAGCACAAGGTCCTGGAGCTCACGGCCGAGAACGAGCGGCTGCAGAAGAAGGTGGAGCAG 1237

Qy    683 cttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatc 742
          ||   ||   |||    |   |  | |||  |  ||||  ||||     ||  |||| ||
Db   1238 CTGTCGCGCGAGCTCAGCACCCTGCGGAACTTGTTCAAGCAGCTGCCCGAGCCCCTGCTC 1297

Qy    743 tc 744
            |
Db   1298 GC 1299


RESULT 13
US-08-012-735-1
; Sequence 1, Application US/08012735
; Patent No. 5360894
;  GENERAL INFORMATION:
;    APPLICANT:  Kishimoto, Tadamitsu
;    APPLICANT:  Hirano, Toshio
;    APPLICANT:  Akira, Shizuo
;    APPLICANT:  Isshiki, Hiroshi
;    APPLICANT:  Tanabe, Osamu
;    APPLICANT:  Kinoshita, Shigemi
;    APPLICANT:  Shimamoto, Takuya
;    TITLE OF INVENTION:  C/EBP2 Gene and Recombinant
;    TITLE OF INVENTION:  C/EBP2
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak &
;      ADDRESSEE:  Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/012,735
;      FILING DATE:  19930203
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/601,094
;      FILING DATE:  22 OCT 1990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1914 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  281..1316
;      OTHER INFORMATION:
US-08-012-735-1


  Query Match             2.1%;  Score 44.4;  DB 1;  Length 1914;
  Best Local Similarity   46.7%;  Pred. No. 0.42;
  Matches  141;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy    443 gacgacacggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaac 502
          |||| || ||    || |  | | |||| |  | | | |  |  |  || |    ||  |
Db    998 GACGCCAAGGCCCCCCCGACCGCCTGCTACGCGGGGGCCGGGCCGGCGCCCTCGCAGGTC 1057

Qy    503 atcaagaggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccag 562

          |  |  | |     ||  |  |||  ||  |  |     ||  |||  |     ||| ||
Db   1058 AAGAGCAAGGCCAAGAAGACCGTGGACAAGCACAGCGACGAGTACAAGATCCGGCGCGAG 1117

Qy    563 gacagcgtcaccaagctgctgacggacgtgcagctgatgaaggggaagcaggagtgcatg 622
            || |  || |    |||  | |  |    |   |   |||  |    |   |   | |
Db   1118 CGCAACAACATCGCCGTGCGCAAGAGCCGCGACAAGGCCAAGATGCGCAACCTGGAGACG 1177

Qy    623 gactccaagctcctggccatgaagcatgagaatgaggctctgtggcgggaggtggccagc 682
          |   |||| ||||||   | | |   ||||| |||   |||  |  | ||||||
Db   1178 CAGCACAAGGTCCTGGAGCTCACGGCCGAGAACGAGCGGCTGCAGAAGAAGGTGGAGCAG 1237

Qy    683 cttcggcagaagcatgcccagcaacagaaagtcgtcaacaagctcattcagttcctgatc 742
          ||   ||   |||    |   |  | |||  |  ||||  ||||     ||  |||| ||
Db   1238 CTGTCGCGCGAGCTCAGCACCCTGCGGAACTTGTTCAAGCAGCTGCCCGAGCCCCTGCTC 1297

Qy    743 tc 744
            |
Db   1298 GC 1299


RESULT 14
US-08-471-033-30
; Sequence 30, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2403 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11..2389
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP3A(a)"
US-08-471-033-30


  Query Match             2.1%;  Score 44.2;  DB 2;  Length 2403;
  Best Local Similarity   44.2%;  Pred. No. 0.48;
  Matches  233;  Conservative    0;  Mismatches  288;  Indels    6;  Gaps    1;

Qy      89 ggcgacggcgttagcccggccctcggcccctctttgcggccgctccctccgcctattccc 148
           | || |||| | |||| |  | |||||  |    ||  ||||    |   |    |   |
Db    1748 GACGGCGGCATCAGCCAGTTCATCGGCGACAAGCTGAAGCCGAAGACCGAGTACGTGATC 1807

Qy     149 tccttgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccg 208
             |    || || ||    |||   |  | |     ||  | |  || |  |  |  |
Db    1808 CAGTACACCGTGAAGGGCAAGCCATCGATTCACCTGAAGGACGAGAACACCGGCTACATC 1867

Qy     209 gccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgc 268
             || |  |  ||        | ||| | |  | ||| |  || | ||  || | ||  |
Db    1868 CACTACGAGGACACCAACAACAACCTGGAGGACTACCAGACCATCAACAAGCGCTTCACC 1927

Qy     269 tggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtg 328
              || |   |  |||  || |  || || ||    || |||||   | | | ||||
Db    1928 ACCGGCACCGACCTGAAGGGCGTGTACCTGATCCTGAAGAGCCAGAACGGCGACGAGGCC 1987

Qy     329 ctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtat 388
             |  | |  ||||||   |    | ||  |     || |  || ||| | |  | |
Db    1988 TGGGGCGACAACTTCATCATCCTGGAGATCAGCCCGAGCGAGAAGCTGCTGAGCCCGGAG 2047

Qy     389 ggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgac 448
             ||   |       | ||| |  |    ||  |     ||        || | | ||
Db    2048 CTGATCAACA------CCAACAACTGGACCAGCACCGGCAGCACCAACATCAGCGGCAAC 2101

Qy     449 acggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaag 508
            ||   |  || | |||   ||  ||    |   || | |||||  |||  ||     ||
Db    2102 ACCCTGACCCTGTACCAGGGCGGCCGCGGCATCCTGAAGCAGAACCTGCAGCTGGACAGC 2161

Qy     509 aggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagc 568
             |       | | |||| |  |     ||| |  |||   ||  | |||     || |
Db    2162 TTCAGCACCTACCGCGTGTACTTCAGCGTGAGCGGCGACGCCAACGTGCGCATCCGCAAC 2221

Qy     569 gtcaccaagctgctgacggacgtgcagctgatgaaggggaagcagga 615
             |  | || |||||   ||   |  |   ||||  ||    ||||
Db    2222 AGCCGCGAGGTGCTGTTCGAGAAGAGGTACATGAGCGGCGCCAAGGA 2268


RESULT  15
US-08-471-044-30
; Sequence 30, Application US/08471044
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2403 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11..2389
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP3A(a)"
US-08-471-044-30


  Query Match             2.1%;  Score 44.2;  DB 3;  Length 2403;
  Best Local Similarity   44.2%;  Pred. No. 0.48;
  Matches  233;  Conservative    0;  Mismatches  288;  Indels    6;  Gaps    1;

Qy      89 ggcgacggcgttagcccggccctcggcccctctttgcggccgctccctccgcctattccc 148
           | || |||| | |||| |  | |||||  |    ||  ||||    |   |    |   |
Db    1748 GACGGCGGCATCAGCCAGTTCATCGGCGACAAGCTGAAGCCGAAGACCGAGTACGTGATC 1807

Qy     149 tccttgctcgagatggatctgcccgtgggccccggcgcggcggggcccagcaacgtcccg 208
             |    || || ||    |||   |  | |     ||  | |  || |  |  |  |
Db    1808 CAGTACACCGTGAAGGGCAAGCCATCGATTCACCTGAAGGACGAGAACACCGGCTACATC 1867

Qy     209 gccttcctgaccaagctgtggaccctcgtgagcgacccggacaccgacgcgctcatctgc 268
             || |  |  ||        | ||| | |  | ||| |  || | ||  || | ||  |
Db    1868 CACTACGAGGACACCAACAACAACCTGGAGGACTACCAGACCATCAACAAGCGCTTCACC 1927

Qy     269 tggagcccgagcgggaacagcttccacgtgttcgaccagggccagtttgccaaggaggtg 328
              || |   |  |||  || |  || || ||    || |||||   | | | ||||
Db    1928 ACCGGCACCGACCTGAAGGGCGTGTACCTGATCCTGAAGAGCCAGAACGGCGACGAGGCC 1987

Qy     329 ctgcccaagtacttcaagcacaacaacatggccagcttcgtgcggcagctcaacatgtat 388
             |  | |  ||||||   |    | ||  |     || |  || ||| | |  | |
Db    1988 TGGGGCGACAACTTCATCATCCTGGAGATCAGCCCGAGCGAGAAGCTGCTGAGCCCGGAG 2047

Qy     389 ggcttccggaaagtggtccacatcgagcagggcggcctggtcaagccagagagagacgac 448
```

```
Db   2048 CTGATCAACA------CCAACAACTGGACCAGCACCGGCAGCACCAACATCAGCGGCAAC 2101

Qy    449 acggagttccagcacccatgcttcctgcgtggccaggagcagctccttgagaacatcaag 508

Db   2102 ACCCTGACCCTGTACCAGGGCGGCCGCGGCATCCTGAAGCAGAACCTGCAGCTGGACAGC 2161

Qy    509 aggaaagtgaccagtgtgtccaccctgaagagtgaagacataaagatccgccaggacagc 568

Db   2162 TTCAGCACCTACCGCGTGTACTTCAGCGTGAGCGGCGACGCCAACGTGCGCATCCGCAAC 2221

Qy    569 gtcaccaagctgctgacggacgtgcagctgatgaaggggaagcagga 615

Db   2222 AGCCGCGAGGTGCTGTTCGAGAAGAGGTACATGAGCGGCGCCAAGGA 2268


Search completed: March 7, 2000, 00:19:53
Job time: 519 sec
```

```
=> fil capl;d que l5; d que l15; s l5 or l15

FILE 'CAPLUS' ENTERED AT 16:25:54 ON 07 MAR 2000
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2000 AMERICAN CHEMICAL SOCIETY (ACS)
```

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications.

```
FILE COVERS 1967 - 7 Mar 2000  VOL 132 ISS 11
FILE LAST UPDATED: 6 Mar 2000  (20000306/ED)

  This file contains CAS Registry Numbers for easy and accurate
  substance identification.

  This file supports REG1stRY for direct browsing and searching of
  all substance data from the REGISTRY file.  Enter HELP FIRST for
  more information.


L5            13 SEA FILE=CAPLUS ABB=ON  HEAT STRESS TRANSCRIPTION FACTOR#


L3           838 SEA FILE=CAPLUS ABB=ON  HSF
L4         24009 SEA FILE=CAPLUS ABB=ON  CHIMER? OR CHIMAER?
L6         16683 SEA FILE=CAPLUS ABB=ON  TRANSCRIPTION FACTORS/CT
L7            46 SEA FILE=CAPLUS ABB=ON  L3(L)L6
L8        266508 SEA FILE=CAPLUS ABB=ON  MUTAT? OR MUTANT?
L9         22556 SEA FILE=CAPLUS ABB=ON  VERTEBRATE#
L10       109920 SEA FILE=CAPLUS ABB=ON  INSECT?
L11       145789 SEA FILE=CAPLUS ABB=ON  MAMMAL?
L12       102198 SEA FILE=CAPLUS ABB=ON  RECOMBINANT
L13       138983 SEA FILE=CAPLUS ABB=ON  CIRCUIT#
L15            4 SEA FILE=CAPLUS ABB=ON  L7 (L) (L4 OR (L8 OR L9 OR L10 OR L11
                 OR L12 OR L13))


L21           17 L5 OR L15

=> fil wpids;d que l16; d que l20

FILE 'WPIDS' ENTERED AT 16:26:07 ON 07 MAR 2000
COPYRIGHT (C) 2000 DERWENT INFORMATION LTD

FILE LAST UPDATED: 01 MAR 2000              <20000301/UP>
>>>UPDATE WEEKS:
MOST RECENT DERWENT WEEK            200011   <200011/DW>
DERWENT WEEK FOR CHEMICAL CODING:   200011
DERWENT WEEK FOR POLYMER INDEXING:  200011
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE

>>> D COST AND SET NOTICE DO NOT REFLECT SUBSCRIBER DISCOUNTS -
                                                SEE HELP COST <<<

>>> FOR UP-TO-DATE INFORMATION ABOUT ALL 'NEW CONTENT' CHANGES TO
```

WPIDS, INCLUDING THE DERWENT CHEMISTRY RESOURCE (DCR),
PLEASE VISIT http://www.derwent.com/newcontent.html <<<

>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES,
SEE http://www.derwent.com/covcodes.html <<<

L16 0 SEA FILE=WPIDS ABB=ON HEAT STRESS TRANSCRIPTION FACTOR#

L4 24009 SEA FILE=CAPLUS ABB=ON CHIMER? OR CHIMAER?
L8 266508 SEA FILE=CAPLUS ABB=ON MUTAT? OR MUTANT?
L9 22556 SEA FILE=CAPLUS ABB=ON VERTEBRATE#
L10 109920 SEA FILE=CAPLUS ABB=ON INSECT?
L11 145789 SEA FILE=CAPLUS ABB=ON MAMMAL?
L12 102198 SEA FILE=CAPLUS ABB=ON RECOMBINANT
L13 138983 SEA FILE=CAPLUS ABB=ON CIRCUIT#
L17 33 SEA FILE=WPIDS ABB=ON HSF
L18 1011857 SEA FILE=WPIDS ABB=ON L4 OR (L8 OR L9 OR L10 OR L11 OR L12 OR L13)
L19 9 SEA FILE=WPIDS ABB=ON L17 AND L18
L20 7 SEA FILE=WPIDS ABB=ON L19 NOT (AC OR TDM)/TI

=> dup rem l21,l20

FILE 'CAPLUS' ENTERED AT 16:26:14 ON 07 MAR 2000
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2000 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'WPIDS' ENTERED AT 16:26:14 ON 07 MAR 2000
COPYRIGHT (C) 2000 DERWENT INFORMATION LTD
PROCESSING COMPLETED FOR L21
PROCESSING COMPLETED FOR L20
L22 24 DUP REM L21 L20 (0 DUPLICATES REMOVED)

=> d ibib ab l22 1-24; fil hom

L22 ANSWER 1 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1999:723189 CAPLUS
DOCUMENT NUMBER: 131:347445
TITLE: Mutant heat shock transcription factor and heat shock promoter for sustained activation of genes by single heat application
INVENTOR(S): Voellmy, Richard
PATENT ASSIGNEE(S): USA
SOURCE: PCT Int. Appl., 51 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9957290	A1	19991111	WO 1999-US9748	19990504

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN,

Searched by Barb O'Bryen, STIC 308-4291

MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 1998-PV84236 19980505

AB The exposure of cells, tissues and organs to "stress", such as elevated temp., stimulates prodn. of active **heat stress transcription factors** (HSF), which in turn, induce expression of genes regulated by stress promoters. Normally, the activity of stress promoters declines after cells, tissues and organs are returned to a normal condition. Mutant forms of HSF, however, can constitutively transactivate stress genes, in the absence of stress. By taking advantage of such mutant HSF, mol. circuits can be devised to provide a sustained expression of a gene of interest using a single application of stress. One form of mol. circuit comprises (a) a first nucleic acid mol. that comprises a gene encoding a transcription factor and a promoter activatable by stress and by the transcription factor, wherein the stress-activatable promoter and the transcription factor gene are operably linked, and (b) a second nucleic acid mol. that comprises a gene of interest and a second promoter activatable by the transcription factor, wherein the second promoter and the gene of interest are operably linked. Thus, HeLa cells transformed with a plasmid contg. the hsp70B promoter fused to human somatotropin cDNA alone, or with this plasmid and a second plasmid contg. the hsp70B promoter fused to a mutant HSF1 gene were subjected to heat stress. In the singly transformed cells, growth hormone prodn. ceased one day after heat treatment. In the doubly transformed cells, prodn. of growth hormone continued for several days after the heat treatment.

L22 ANSWER 2 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1999-418930 [35] WPIDS
DOC. NO. CPI: C1999-123171
TITLE: New isolated Toxoplasma gondii nucleic acids used, e.g. to treat infection caused by this microorganism.
DERWENT CLASS: B04 D16
INVENTOR(S): LUTZ, S B; MILHAUSEN, M J; NG, R K
PATENT ASSIGNEE(S): (HESK-N) HESKA CORP
COUNTRY COUNT: 84
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9932633	A1	19990701	(199935)*	EN	380

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

AU 9919348 A 19990712 (199950)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9932633	A1	WO 1998-US27137	19981218
AU 9919348	A	AU 1999-19348	19981218

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9919348	A	Based on	WO 9932633

PRIORITY APPLN. INFO: US 1997-994825 19971219

AB WO 9932633 A UPAB: 19990902

NOVELTY - Isolated Toxoplasma gondii nucleic acids and polypeptides are new.

DETAILED DESCRIPTION - A novel isolated nucleic acid molecule (NAM) encodes an immunogenic Toxoplasma gondii (TG) protein that can be identified by a method comprising:

(a) immunoscreening a library selected from a TG genomic expression library and a TG cDNA expression library with an antiserum, where the antiserum is selected from antiserum: raised against TG oocysts, raised against TG bradyzoites, raised against TG infected cat gut and isolated from a cat immune to TG infection; and

(b) identifying a nucleic acid molecule in the library that expresses a protein that selectively binds to an antibody in the antiserum.

INDEPENDENT CLAIMS are also included for the following:

(1) isolating a NAM encoding an immunogenic TG protein comprising:

(a) as in (Aa) and (Ab); and

(b) recovering the NAM from the library;

(2) an isolated immunogenic TG protein that can be identified by a method comprising:

(a) as in (Aa) and (Ab), and

(b) identifying a protein expressed from the library that selectively binds to antibodies in the antiserum;

(3) an isolated NAM that hybridizes under stringent hybridization conditions with a gene comprising a nucleic acid sequence selected from the 83 sequences (given in the specification), e.g. T. gondi DNA sequences of 357 and 339 base pairs;

(4) an isolated NAM that hybridizes under stringent hybridization conditions with a gene comprising a NAM selected from nTG2339, nTG4526, nTG41478, nTG5657, nTG51029, nTG6425, nTG7417, nTG8507, nTG9718, nTG10441, nTG11428, nTG13282, nTG15304, nTG16284, nTG17690, nTG18313, nTG19389, nTG21548, nTG22310, nTG23220, nTG24642, nTG25381, nTG26432, nTG27282, nTG466, nTGGG30539, nTG311233, nTG32411, nTG33441, nTG34491, nTG35387, nTG36417, nTG37416, nTG38500, nTG40321, nTG41513, nTG42528, nTG43375, nTG44543, nTG45573, nTG461835, nTG48604, nTG48549, nTG49270, nTG50306, nTG51804, nTG52867, nTG531434, nTG54680, nTG55296, nTG56723, nTG57270, nTG58503, nTG60322, nTG61390, nTG62699, nTG63419, nTG64303, nTG65696, nTG66173, nTG67369, nTG68566, nTG69616, nTG70762, nTG71236, nTG72569, nTG73232, nTG74276, nTG75309, nTG76534, nTG76423, nTG77327, nTG78444 and nTG79928;

(5) an isolated immunogenic protein encoded by a NAM that hybridizes under stringent hybridization conditions with a gene comprising the complement of a nucleic acid sequence selected from sequences as in (3);

(6) an isolated immunogenic protein encoded by a NAM that hybridizes under stringent conditions with a gene comprising a NAM selected from NAMs as in (4);

(7) an isolated antibody that selectively binds to a protein as in (2), (4) or (5);

(8) detecting parasite cysts or oocysts in feces comprising:

(a) contacting a sample of feces with a solid support capable of binding oocysts;

(b) allowing the sample to dry onto the solid support;

(c) washing the sample on the solid support with an aqueous wash solution;

(d) adding an aqueous elution solution to the sample and eluting DNA from the sample into the aqueous elution solution by heating;

(e) PCR amplifying oocyst-specific DNA with primers specific to the oocyst being detected; and

(f) detecting the presence of a PCR amplification product resulting from amplification of oocyst-specific DNA in the sample where the presence of the product indicates the presence of cysts or oocysts in the feces;

(9) a **recombinant** molecule comprising a NAM as in (A), (3) or (4), operatively linked to a transcription control sequence; and

(10) a **recombinant** virus or cell comprising a NAM as in (A), (3) or (4).

USE - The TG NAMS, immunogenic proteins and antibodies to the proteins can be used to inhibit TG oocyst shedding in a cat due to infection with TG (claimed). They can be used for preventing TG infection and for preventing the spread of TG infection. They can also be used for detecting TG infection. The detection method can be used to detect parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts such as Cryptosporidium oocysts and Toxoplasma oocysts.

Dwg.0/0

L22 ANSWER 3 OF 24 CAPLUS COPYRIGHT 2000 ACS

ACCESSION NUMBER: 1999:637755 CAPLUS
DOCUMENT NUMBER: 132:19414
TITLE: The mammalian HSF4 gene generates both an activator and a repressor of heat shock genes by alternative splicing
AUTHOR(S): Tanabe, Masako; Sasai, Noriaki; Nagata, Kazuhiro; Liu, Xiao-Dong; Liu, Phillip C. C.; Thiele, Dennis J.; Nakai, Akira
CORPORATE SOURCE: Department of Molecular and Cell Biology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, 606-8397, Japan
SOURCE: J. Biol. Chem. (1999), 274(39), 27845-27856
CODEN: JBCHA3; ISSN: 0021-9258
PUBLISHER: American Society for Biochemistry and Molecular Biology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The expression of heat shock genes is controlled at the level of transcription by members of the heat shock transcription factor family in vertebrates. HSF4 is a mammalian factor characterized by its lack of a suppression domain that modulates formation of DNA-binding homotrimer. Here, we have detd. the exon structure of the human HSF4 gene and identified a major new isoform, HSF4b, derived by alternative RNA splicing events, in addn. to a previously reported HSF4a isoform. In mouse tissues HSF4b mRNA was more abundant than HSF4a as examd. by reverse transcription-polymerase chain reaction, and its protein was detected in the brain and lung. Although both mouse HSF4a and HSF4b form trimers in the absence of stress, these two isoforms exhibit different transcriptional activity; HSF4a acts as an inhibitor of the constitutive expression of heat shock genes, and hHSF4b acts as a transcriptional activator. Furthermore HSF4b but not HSF4a complements the viability defect of yeast cells lacking HSF. Moreover, heat shock and other stresses stimulate transcription of target genes by HSF4b in both yeast and mammalian cells. These results suggest that differential splicing of HSF4 mRNA gives rise to both an inhibitor and activator of tissue-sp. heat shock gene expression.

L22 ANSWER 4 OF 24 CAPLUS COPYRIGHT 2000 ACS

ACCESSION NUMBER: 1999:429403 CAPLUS
DOCUMENT NUMBER: 131:195949
TITLE: Identification of a key residue in Drosophila heat shock factor-DNA interaction by analytical ultracentrifugation
AUTHOR(S): Park, Jinku; Kim, Seha; Kim, Soon-Jong
CORPORATE SOURCE: Department of Chemistry, Mokpo National University,

Searched by Barb O'Bryen, STIC 308-4291

Muan, 534-729, S. Korea
SOURCE: Bull. Korean Chem. Soc. (1999), 20(6), 636-638
CODEN: BKCSDE; ISSN: 0253-2964
PUBLISHER: Korean Chemical Society
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The authors generated two mutant heat shock factor DNA binding domains where arginine 102 and asparagine 105 on the DNA recognition helix where replaced with alanines. The effect of mutation on protein-DNA interaction were studied. The highly conserved arginine appears to play a major role in heat shock element recognition.

L22 ANSWER 5 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1999:574619 CAPLUS
DOCUMENT NUMBER: 131:281526
TITLE: Increased resistance of the radiosensitive M10 mutant cells of the L5178y mouse lymphoma cell line to heat-induced apoptosis
AUTHOR(S): Ostapenko, Valentina V.; Wang, Xinjiang; Ohnishi, Ken; Takahashi, Akihisa; Yamamoto, Itsuo; Tanaka, Yoshimasa; Ohnishi, Takeo
CORPORATE SOURCE: Department of Radiology, Kansai Medical University, Osaka, 570-8506, Japan
SOURCE: Radiat. Res. (1999), 152(3), 321-327
CODEN: RAREAE; ISSN: 0033-7587
PUBLISHER: Radiation Research Society
DOCUMENT TYPE: Journal
LANGUAGE: English
AB M10 cells, which are deficient in the repair of DNA DSBs and are therefore radiosensitive, are about twofold more thermoresistant than their parental L5178Y cells. We found that, after heat shock at 43.degree.C under conditions resulting in 10% survival (D10), M10 cells did not undergo apoptosis, whereas L5178Y cells did undergo apoptosis. M10 cells, but not L5178Y cells, constitutively expressed Hsp72 protein. Moreover, the M10 cells accumulated higher amts. of the heat-inducible form of Hsp72. The patterns of activation of the DNA-binding activity of HSF (heat-shock factor) differed in M10 and L5178Y cells. In response to heat shock, M10 cells accumulated greater amts. of Trp53 protein (formerly known as p53) than the parental cells. Cdkn1a (formerly known as p21, Waf1) was constitutively expressed and further accumulated after heat shock only in M10 cells. We suggest that heat-inducible Hsp72 to a larger extent, and constitutive Hsp72 to a lesser extent, together with Cdkn1a may be involved in the protection of M10 cells against heat-induced apoptosis. Apoptosis in these cells is likely to occur in Trp53-dependent manner.

L22 ANSWER 6 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1999:424045 CAPLUS
DOCUMENT NUMBER: 131:183295
TITLE: Differential activation of some transcription factors during rat liver ischemia, reperfusion, and heat shock
AUTHOR(S): Tacchini, Lorenza; Radice, Laura; Bernelli-Zazzera, Aldo
CORPORATE SOURCE: Istituto di Patologia Generale dell'Universita degli Studi di Milano, Centro di Studio sulla Patologia Cellulare del CNR, Milan, 20133, Italy
SOURCE: J. Cell. Physiol. (1999), 180(2), 255-262
CODEN: JCLLAX; ISSN: 0021-9541
PUBLISHER: Wiley-Liss, Inc.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Cells respond to external stimuli by changes in gene expression that are largely dependent on transcription factors (TFs). We studied the behavior

of some TFs in rat liver during ischemia, postischemic reperfusion, and heat shock. Knowledge of the conditions at the end of ischemia is essential to understand changes occurring at reperfusion. The TFs investigated are known to be typically responsive to heat shock (HSF), hypoxia (HIF-1), pro- and antioxidant conditions (AP-1), or to various environmental changes (HNF-1 and ATF/CREB family). The most relevant new information includes the following: (1) Liver ischemia activates extremely rapidly the DNA binding capacity of HSF, soon followed by analogous activation of HIF-1 and AP-1. (2) After a certain lag time from the activation of HIF-1, mRNAs accumulate for two glycolytic enzymes, in particular Aldolase A and Heme Oxygenase 1, which contain HIF-1 sequences in their promoters. (3) Reperfusion, which is known to further increase the binding of HSF and to induce NF.kappa.B binding, abrogates or decreases the binding of HIF-1 and AP-1, stimulated by ischemia, and activates the binding of ATF/CREB. Later on, a second peak of AP-1 binding is induced. (4) Heat shock activates both ischemia-responsive and reperfusion-responsive TFs. (5) Preliminary expts. of supergelshift reveal that the activation of AP-1 at reperfusion or upon heat shock may result from the different involvement of the component subunits.

L22 ANSWER 7 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1999:148716 CAPLUS
DOCUMENT NUMBER: 131:1362
TITLE: Mutations in the Schizosaccharomyces pombe heat shock factor that differentially affect responses to heat and cadmium stress
AUTHOR(S): Saltsman, K. A.; Prentice, H. L.; Kingston, R. E.
CORPORATE SOURCE: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, 02114, USA
SOURCE: Mol. Gen. Genet. (1999), 261(1), 161-169
CODEN: MGGEAE; ISSN: 0026-8925
PUBLISHER: Springer-Verlag
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Heat shock factor (hsf) is the transcriptional activator that governs the transcriptional response of eukaryotic cells to stressful conditions. The structure and regulation of hsf is highly conserved. We describe deletion mutations in hsf+ that alter the ability of Schizosaccharomyces pombe to respond to different stressful conditions. One mutation causes increased sensitivity to cadmium while maintaining near normal sensitivity to heat stress, while another mutation confers increased sensitivity to heat stress but retains normal sensitivity to cadmium. Despite the differential sensitivity of these two strains to cadmium and heat stress, the mutant hsf proteins in each strain were activated by both cadmium and heat. However, we found that these mutations differentially affected the ability of hsf to activate different promoters: one mutated hsf activated the ssp1+ gene better than the wis2+ gene following either stress, while the other mutated hsf activated wis2+ better than ssp1+. We propose that the differential ability of strains that contain these mutant hsfs to survive cadmium and heat stress is not caused by differences in activation of hsf, but is caused instead by differential abilities of the mutant hsfs to activate the appropriate sets of genes needed for survival.

L22 ANSWER 8 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-414102 [35] WPIDS
DOC. NO. CPI: C1998-125053
TITLE: Method for modulating synthesis of heat-shock protein - by administering **mutant** heat shock transcription factors, used, e.g to protect cells against chemotherapy.
DERWENT CLASS: B04 D16
INVENTOR(S): VOELLMY, R W
Searched by Barb O'Bryen, STIC 308-4291

PATENT ASSIGNEE(S): (UYMI-N) UNIV MIAMI
COUNTRY COUNT: 81
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9831803	A1	19980723	(199835)*	EN	84

RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

AU 9860320	A	19980807	(199901)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9831803	A1	WO 1998-US1038	19980121
AU 9860320	A	AU 1998-60320	19980121

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9860320	A	Based on	WO 9831803

PRIORITY APPLN. INFO: US 1997-914646 19970819; US 1997-35662 19970121

AB WO 9831803 A UPAB: 19980904
Synthesis of heat-shock proteins (I) is increased in a cell by introducing a positive-acting **mutant** mutHSF (II; **mutated** heat shock transcription factor), particularly to induce a protected state in the cell. Also new is inducing a sensitised state in a cell, or inhibiting stress-induced (I) synthesis, by introducing a negative-acting mut **HSF** (III).

USE - (II), or nucleic acid encoding it or cells expressing it, are administered to a subject: (i) to protect cells against damage caused by chemotherapeutics, UV-B light, sepsis, hyperthermia, inflammatory responses or cytokines, oxidative stress and ischaemia, particularly to increase resistance of normal cells to anti-tumour agents, or (ii) to increase immunogenicity of cancer cells. (III) are used to render cancer cells more sensitive to killing by heat and/or other stresses, and also to induce apoptosis.

ADVANTAGE - (II) are active in absence of stress, unlike wild-type HSF1, even when over expressed, and eliminate the need for cytotoxic agents for regulating the heat-shock system.

Dwg.0/3

L22 ANSWER 9 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-427558 [36] WPIDS
DOC. NO. CPI: C1998-128880
TITLE: Isolated haematopoietic signalling factor - used to develop products for diagnosis and treatment of, e.g. haematopoietic disorders such as leukaemia and acute or chronic inflammation.
DERWENT CLASS: B04 D16
INVENTOR(S): RUBEN, S M; SOPPET, D R
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 81
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9831792	A1	19980723	(199836)*	EN	71
RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
AU 9859210	A	19980807	(199901)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9831792	A1	WO 1998-US854	19980116
AU 9859210	A	AU 1998-59210	19980116

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9859210	A	Based on	WO 9831792

PRIORITY APPLN. INFO: US 1997-35577 19970116

AB WO 9831792 A UPAB: 19981008
The following are claimed (1) an isolated nucleic acid molecule (I) comprises a polynucleotide at least 95% identical to a nucleotide sequence selected from: (a) NS encoding a haematopoietic signalling factor (**HSF**) polypeptide having an amino acid sequence at positions from -26 to 353 of 379 aa sequence (S1) (given in the specification); (b) NS encoding a **HSF** polypeptide having an amino acid sequence at positions from -25 to 353 of (S1); (c) NS encoding an amino acid sequence at positions from 1 to 353 of (S1); (d) NS encoding a **HSF** polypeptide having the complete amino acid sequence encoded by a cDNA clone contained in ATCC 97731; (e) a NS encoding a mature **HSF** polypeptide having an amino acid sequence encoded by a cDNA clone contained in ATCC 97731, and (f) a NS complementary to any of the NSs in (a)-(e); (2) an isolated NAM comprising a PN which hybridises under stringent hybridisation conditions to a PN having a NS as in (1a)-(1e), where the PN which hybridises does not hybridise under stringent hybridisation conditions to a PN having a NS consisting of only A residues or of only T residues; (3) an isolated NAM comprising a PN which encodes an amino acid sequence of an epitope-bearing portion of a **HSF** polypeptide having an amino acid sequence as in (A) (a)-(e); (4) an isolated NAM comprising PN having a sequence at least 95% identical to a sequence selected from: (a) a NS of a fragment of a 1545 bp sequence (S2) (given in the specification), where the fragment comprises at least 50 contiguous nucleotides of (S2), provided that the NS is not 418, 443, 513, 352, or 449 bp sequence (all sequences are given in the specification), or any subfragment, and (b) NS complementary to a nucleotide sequence in (a); (5) a method for preparation of a **recombinant** vector comprising inserting an isolated (I) into a vector; (6) a **recombinant** vector produced by the method of (5); (7) a method of preparing a **recombinant** host cell comprising introducing a **recombinant** vector as in (6) into a host cell; (8) a **recombinant** host cell produced by the method as in (7); (9) an isolated **HSF** polypeptide similar to (1); (10) an isolated NAM comprising a PN encoding an **HSF** polypeptide where, except for 1 to 50 conservative amino acid substitutions, the polypeptide has a sequence selected from (a)-(f) as in (I), and (11) an isolated **HSF** polypeptide where, except

for 1 to 50 conservative amino acid substitutions, the polypeptide has a sequence selected from (a)-(f) as in (I); (L) an isolated antibody that binds specifically to a **HSF** polypeptide as in (I).

USE - The **HSF** polypeptides can be used for modulating haematopoietic activities. The products can be used for treating e.g. leukaemias, lymphomas, Hodgkin's disease or non-Hodgkin's lymphomas. They can also be used for treating acute or chronic inflammation. They can also be used for modulating the development of haematopoietic stem cells. The products can also be used for detection, diagnosis and drug screening. Dwg.0/3

L22 ANSWER 10 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1998:746752 CAPLUS
DOCUMENT NUMBER: 130:135407
TITLE: Heat stress response and **heat stress transcription factors**
AUTHOR(S): Scharf, Klaus-Dieter; Hohfeld, Ingo; Nover, Lutz
CORPORATE SOURCE: Molecular Cell Biology, Biocenter, Goethe-University-Frankfurt, Frankfurt/Main, D-60439, Germany
SOURCE: J. Biosci. (Bangalore, India) (1998), 23(4), 313-329
CODEN: JOBSDN; ISSN: 0250-5991
PUBLISHER: Indian Academy of Sciences
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 171 refs. Expression of heat shock protein (HSP)-coding genes is controlled by **heat stress transcription factors** (Hsfs). They are structurally and functionally conserved throughout the eukaryotic kingdom. In addn. to the DNA-binding domain with the helix-turn-helix motif essential for DNA recognition, three functional parts in the C-terminal activator domain were characterized: (i) the HR-A/B region is responsible for oligomerization and activity control, (ii) the nuclear localizing signal (NLS) formed by a cluster of basic amino acid residues which is required and sufficient for nuclear import and (iii) short C-terminal peptide motifs with a central Trp residue (AHA elements). These three parts are indispensible for the activator function. A peculiarity of plants is the heat shock-inducible new synthesis of Hsfs. In tomato HsfA1 is constitutively expressed, whereas Hsfs A2 and B1 are heat shock-inducible proteins themselves. We used Hsf knock-out strains of yeast and transient reporter assays in tobacco protoplasts for functional anal. of Hsf-coding cDNA clones and mutants derived from them. HsfA2, which in tomato cell cultures is expressed only after heat shock induction, tends to form large cytoplasmic aggregates together with other HSPs (heat stress granules). In the transient expression assay its relatively low activator potential is evidently due to the inefficient nuclear import. However, the intramol. shielding of the NLS can be released either by deletion of a short C-terminal fragment or by coexpression with HsfA1, which forms hetero-oligomers with HsfA2.

L22 ANSWER 11 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1998:233176 CAPLUS
DOCUMENT NUMBER: 128:292793
TITLE: Molecular and applied aspects of the heat stress response and of common stress tolerance in plants
AUTHOR(S): Schoffl, F.; Prandl, R.; Hinderhofer, K.; Reindl, A.
CORPORATE SOURCE: Universitat Tubingen, Lehrstuhl Allgemeine Genetik, Tubingen, D-72076, Germany
SOURCE: Acta Physiol. Plant. (1997), 19(4), 549-550
CODEN: APPLDE; ISSN: 0137-5881
PUBLISHER: Polish Academy of Sciences
DOCUMENT TYPE: Journal

LANGUAGE: English
AB Data from research on Arabidopsis thaliana suggest that a neg. control mechanism regulates the activity of certain heat-shock transcription factors. This can be overcome by heat shock in wild type or overexpression in transgenic plants.

L22 ANSWER 12 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1997:553932 CAPLUS
DOCUMENT NUMBER: 127:258501
TITLE: **Heat stress transcription factors** from tomato can functionally replace HSF1 in the yeast Saccharomyces cerevisiae
AUTHOR(S): Boscheinen, O.; Lyck, R.; Queitsch, C.; Treuter, E.; Zimarino, V.; Scharf, K.-D.
CORPORATE SOURCE: Molecular Cell Biology, Biocenter of the J. W. Goethe University, Frankfurt, D-60439, Germany
SOURCE: Mol. Gen. Genet. (1997), 255(3), 322-331
CODEN: MGGEAE; ISSN: 0026-8925
PUBLISHER: Springer
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The fact that yeast HSF1 is essential for survival under nonstress conditions can be used to test heterologous Hsfs for the ability to substitute for the endogenous protein. Our results demonstrate that like Hsf of Drosophila, tomato Hsfs A1 and A2 can functionally replace the corresponding yeast protein, but Hsf B1 cannot. In addn. to survival at 28.degree.C, we checked the transformed yeast strains for temp. sensitivity of growth, induced thermotolerance and activator function using two different lacZ reporter constructs. Tests with full-length Hsfs were supplemented by assays suing mutant Hsfs lacking parts of their C-terminal activator region or oligomerization domain, or contg. amino acid substitutions in the DNA-binding domain. Remarkably, results with the yeast system are basically similar to those obtained by the anal. of the same Hsfs as transcriptional activators in a tobacco protoplast assay. Most surprising is the failure of HsfB1 to substitute for the yeast Hsf. The defect can be overcome by addn. to HsfB1 of a short C-terminal peptide motif from HsfA2 (34 amino acid residues), which represents a type of minimal activator necessary for interaction with the yeast transcription app. Deletion of the oligomerization domain (HR-A/B) does not interfere with Hsf function for survival or growth at higher temps. But monomeric Hsf has a markedly reduced affinity for DNA, as shown by lacZ reporter and band-shift assays.

L22 ANSWER 13 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1997:318124 CAPLUS
DOCUMENT NUMBER: 127:76959
TITLE: Intracellular distribution and identification of the nuclear localization signals of two plant **heat -stress transcription factors**
AUTHOR(S): Lyck, Ruth; Harmening, Uwe; Hohfeld, Ingo; Treuter, Eckardt; Scharf, Klaus Dieter; Nover, Lutz
CORPORATE SOURCE: Biocenter, J. W. Goethe-University, Frankfurt/Main, D-60439, Germany
SOURCE: Planta (1997), 202(1), 117-125
CODEN: PLANAB; ISSN: 0032-0935
PUBLISHER: Springer
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Similar to **heat-stress transcription factors** (HSFs) from non-plant sources, HSFA1 and HSFA2 from tomato (Lycopersicon esculentum) contain 2 conserved clusters of basic amino acid

residues (K/R1 and K/R2) which might serve as nuclear localization signal (NLS) motifs. Mutation of either one of them and functional testing of the corresponding proteins in a transient expression assay using tobacco (Nicotiana plumbaginifolia) protoplasts gave the following results. Whereas K/R1, positioned in all HSFs at the C-terminus of the DNA-binding domain, had no influence on nuclear import, the K/R1 mutants were impaired in their interaction with the DNA (band-shift assays). In contrast to this, mutants of the K/R2 motif, found 15-20 amino acid residues C-terminal of the oligomerization domain (HR-A/B region), had wild-type activity in DNA-binding but were defective in nuclear import. Thus, for the related tomato HSFA1 and HSFA2 the K/R2 cluster represents the only NLS motif, and in this function it cannot be replaced by K/R1.

L22 ANSWER 14 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1997:168255 CAPLUS
DOCUMENT NUMBER: 127:77416
TITLE: Developmental control of heat shock and chaperone gene expression. Part 1. Plants and nonmammals. Heat stress proteins and transcription factors
AUTHOR(S): Nover, L.; Scharf, K. D.
CORPORATE SOURCE: Biocenter, Johann-Wolfgang-Goethe-Univ., Frankfurt/Main, D-60439, Germany
SOURCE: Cell. Mol. Life Sci. (1997), 53(1), 80-103
CODEN: CMLSFI; ISSN: 1420-682X
PUBLISHER: Birkhaeuser
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 434 refs. is given on heat stress proteins (HSP) as part of interacting chaperone systems, the HSP100 family, the HSP90 system, the HSP70/DnaK chaperone system, the HSP69/GroEL chaperone system, the HSP20 family, stress proteins as components of proteolytic systems, and peptidyl-prolyl cis/trans isomerases. **Heat stress transcription factors** (HSF) are described, their basic structure, the DNA-binding domain, heptad hydrophobic repeats, nuclear localization signal, and the C-terminal activator domain. The developmental control of heat stress gene expression in plants is described.

L22 ANSWER 15 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1997:63539 CAPLUS
DOCUMENT NUMBER: 126:142186
TITLE: Distinct stress-inducible and developmentally regulated heat shock transcription factors in Xenopus oocytes
AUTHOR(S): Gordon, Sandra; Bharadwaj, Steve; Hnatov, Alex; Ali, Adnan; Ovsenek, Nick
CORPORATE SOURCE: Dep. of Anatomy and Cell Biology, Univ. of Saskatchewan, Saskatchewan, S7N 5E5, Can.
SOURCE: Dev. Biol. (1997), 181(1), 47-63
CODEN: DEBIAO; ISSN: 0012-1606
PUBLISHER: Academic
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The presence of a maternal pool of heat shock factor (HSF) in Xenopus oocytes has been suggested by two lines of evidence from previous studies. First, heat shock response element (HSE)-binding activity is induced in heat-shocked eggs and embryos prior to expression of zygotic HSF. Second, expression from microinjected heat shock protein promoters in oocytes is induced upon heat shock. To date, however, endogenous oocyte HSF mols. have not been detected, nor has induction of HSE-binding activity been directly demonstrated. Here we report the detection of distinct stress-inducible and developmentally regulated HSE-binding activities of

endogenous oocyte factors. Exposure of defolliculated oocytes to heat, cadmium, and arsenite resulted in the formation of tan HSE-specific complex detectable by gel mobility shift assay. Induction of HSE-binding activity by each of these stressors corresponded to increased expression from a microinjected hsp70 promoter. The stress-inducible HSE-binding complex was recognized by antiserum against mammalian HSF1, but not by HSF2 antiserum, suggesting that a Xenopus homolog of HSF1 is the major component of this activity. The HSE-binding activity of HSF1 was induced by stress treatments of stage I through VI oocytes, an indication that it is responsive to stress throughout oogenesis. During recovery from heat shock, the HSF1-HSE complex rapidly declined to control levels, but was induced for prolonged periods in oocytes exposed to continuous stress, a pattern unlike the transient activation previously obsd. in fertilized eggs or embryos. The kinetics of HSF1 activation in oocytes suggests that a key protein(s) regulating attenuation of the stress response is present at exceedingly low levels or is somehow modified during preembryonic development. We also detected an unusual constitutive HSE-binding complex in unstressed stage I and II oocytes, but not in later stage oocytes, eggs, developing embryos, or A6 cells. This constitutive complex was unaffected by heat or chem. treatments and was not recognized by either HSF1 or HSF2 antiserum. Appearance of the constitutive HSE-binding activity during oogenesis corresponding closely with peak levels of hsp70 mRNA detected by Northern blot anal. of RNA from staged oocytes. We suggest that the constitutive HSE-binding activity in early oocytes is formed by a unique developmentally regulated heat shock factor that may play a role in the expression of heat shock proteins during early stages of oogenesis.

L22 ANSWER 16 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1996:197580 CAPLUS
DOCUMENT NUMBER: 124:280908
TITLE: Solution structure of the DNA-binding domain of the tomato **heat-stress transcription factor** HSF24
AUTHOR(S): Schultheiss, Juergen; Kunert, Olaf; Gase, Uwe; Scharf, Klaus-Dieter; Nover, Lutz; Rueterjans, Heinz
CORPORATE SOURCE: Dep. Biophysical Chem., Goethe-Universitaet, Frankfurt, Germany
SOURCE: Eur. J. Biochem. (1996), 236(3), 911-21
CODEN: EJBCAI; ISSN: 0014-2956
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Two-dimensional-NMR and 3-dimensional-NMR expts. were performed to det. the soln. structure of the DNA-binding domain of the tomato **heat-stress transcription factor** HSF24. Samples of uniformly 15N-labeled and 15N,13C-labeled recombinant proteins were used in the investigation. A near-complete assignment of the backbone 1H, 15N, and 13C resonances was obtained by 3-dimensional triple-resonance expts., whereas 3-dimensional 15N-TOCSY-heteronuclear-single-quantum-correlation-spectroscopy, HCCH-COSY and HCCH-TOCSY spectra were recorded for side-chain assignments. 885 Non-redundant distance constraints from 2-dimensional-homonuclear and 3-dimensional-15N-edited and 13C-edited NOESY spectra and 40 hydrogen-bond constraints from exchange expts. were used for structure calcns. The resulting 3-dimensional structure contains a 3-helix bundle and a small 4-stranded antiparallel .beta.-sheet that forms a hydrophobic core. The 2 C-terminal helixes are parts of a highly conserved helix-turn-helix motif that is probably involved in DNA recognition and binding. In contrast to heat-stress factors from yeast and animals, the plant heat-stress factors lack a loop of 11 amino acid residues inserted between .beta.3 and .beta.4. This leads to a tight turn between these .beta.-strands.

L22 ANSWER 17 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1997:58793 CAPLUS
DOCUMENT NUMBER: 126:99885
TITLE: The Hsf world: Classification and properties of plant **heat stress transcription factors**
AUTHOR(S): Nover, Lutz; Scharf, Klaus-Dieter; Gagliardi, Dominique; Vergne, Philipe; Czarnecka-Verner, Eva; Gurley, William B.
CORPORATE SOURCE: Biocenter the Goethe University, Frankfurt/Main, D-60439, Germany
SOURCE: Cell Stress Chaperones (1996), 1(4), 215-223
CODEN: CSCHFG; ISSN: 1355-8145
PUBLISHER: Churchill Livingstone
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review and discussion with 58 refs. on the classification and properties of **heat stress transcription factors** of plants. Based on the partial or complete sequences of 14 plant **heat stress transcription factors** (Hsfs) from tomato, soybean, Arabidopsis and maize the authors propose a general nomenclature with two basic classes, i.e. classes A and B each contg. two or more types of Hsfs (HsfA1, HsfA2 etc.). Despite some plant-specific peculiarities, essential functional domains and modules of these proteins are conserved among plants, yeast, Drosophila and vertebrates. A revised terminol. of these parts follows recommendations agreed upon among the authors and representatives from other labs. working in this field. Similar to the situation with the small heat shock proteins (sHsps), the complexity of the hsf gene family in plants appears to be higher than in other eukaryotic organisms.

L22 ANSWER 18 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1994:24570 CAPLUS
DOCUMENT NUMBER: 120:24570
TITLE: Two cDNAs for tomato **heat stress transcription factors**
AUTHOR(S): Scharf, Klaus Dieter; Rose, Sonja; Thierfelder, Joerg; Nover, Lutz
CORPORATE SOURCE: Inst. Plant Biochem., Halle, O-4050, Germany
SOURCE: Plant Physiol. (1993), 102(4), 1355-6
CODEN: PLPHAY; ISSN: 0032-0889
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Southwestern screening of a .lambda.gt11 cDNA expression library of tomato (Lycopersicon peruvianum) resulted in the isolation of 3 different heat stress factor (HSF) clones. The cDNA sequences of hsf8 and hsf30 and some structural features of the corresponding proteins are reported.

L22 ANSWER 19 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1994:24870 CAPLUS
DOCUMENT NUMBER: 120:24870
TITLE: Promoter specificity and deletion analysis of three **heat stress transcription factors** of tomato
AUTHOR(S): Treuter, Eckhardt; Nover, Lutz; Ohme, Karin; Scharf, Klaus Dieter
CORPORATE SOURCE: Inst. Plant Biochem., Halle-Salle, 06120, Germany
SOURCE: Mol. Gen. Genet. (1993), 240(1), 113-25
CODEN: MGGEAE; ISSN: 0026-8925
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Transient expression assays in transformed tobacco (Nicotiana

plumbaginifolia) mesophyll protoplasts were used to test the activity of three tomato **heat stress transcription factors**, HSF24, HSF8 and HSF30, in a trans-activation and a trans-repression assay. The results document differences between the three HSFs with respect to their response to the configuration of heat stress promoter elements (HSEs) in the reporter construct (promoter specificity) and to the stress regime used for activation. Anal. of C-terminal deletions identified acidic sequence elements with a central tryptophan residue, which are important for HSF activity control. Surprisingly, heterologous HSFs from Drosophila and human cells, but not from yeast, were also functional as heat stress-induced transcription factors in this tobacco protoplast system.

L22 ANSWER 20 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1992-217013 [26] WPIDS
DOC. NO. CPI: C1992-098273
TITLE: DNA fragment encoding Drosophila or human heat shock factor protein - and use of corresp. monoclonal antibodies for diagnosing abnormal stress conditions in cells.
DERWENT CLASS: B04 D16
INVENTOR(S): CLOS, J; RABINDRAN, S; WESTWOOD, J T; WU, C
PATENT ASSIGNEE(S): (USDC) US DEPT OF COMMERCE; (USSH) US DEPT HEALTH & HUMAN SERVICES
COUNTRY COUNT: 18
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9209617	A1	19920611	(199226)*	EN	75
RW: AT BE CH DE DK ES FR GB GR IT LU NL SE					
W: AU CA JP					
AU 9190723	A	19920625	(199239)		
EP 559770	A1	19930915	(199337)	EN	
R: AT BE CH DE DK ES FR GB GR IT LI LU NL SE					
JP 06502540	W	19940324	(199417)		25
AU 656350	B	19950202	(199513)		
EP 559770	A4	19950405	(199613)		
US 5756343	A	19980526	(199828)		
JP 2845623	B2	19990113	(199907)		42

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 9209617	A1		WO 1991-US8592	19911122
AU 9190723	A		AU 1991-90723	19911122
			WO 1991-US8592	19911122
EP 559770	A1		WO 1991-US8592	19911122
			EP 1992-901003	19911122
JP 06502540	W		WO 1991-US8592	19911122
			JP 1992-501958	19911122
AU 656350	B		AU 1991-90723	19911122
EP 559770	A4		EP 1992-901003	
US 5756343	A	Div ex	US 1990-617910	19901126
			US 1994-178477	19940107
JP 2845623	B2		WO 1991-US8592	19911122
			JP 1992-501958	19911122

FILING DETAILS:

PATENT NO	KIND	PATENT NO

AU 9190723	A	Based on	WO 9209617
EP 559770	A1	Based on	WO 9209617
JP 06502540	W	Based on	WO 9209617
AU 656350	B	Previous Publ.	AU 9190723
		Based on	WO 9209617
JP 2845623	B2	Previous Publ.	JP 06502540
		Based on	WO 9209617

PRIORITY APPLN. INFO: US 1990-617910 19901126; US 1994-178477 19940107

AB WO 9209617 A UPAB: 19931006
A DNA fragment (I) encoding a Drosophila heat shock factor (**HSF**) protein or fragment (sequence given in specification) is new. Also new are: (a) a DNA fragment (II) encoding a human **HSF** protein or fragment (sequence given in specification); (b) a **recombinant** DNA molecule comprising (I) or (II), and a vector for introducing the molecule into prokaryotic or eukaryotic cells; (c) a host cell stably transformed or transfected with (b) for expression of **HSF**; (d) purified antibodies specific for Drosophila or human **HSF**; (e) a method for the detection of human **HSF** comprising contacting a reagent which specifically reacts with the protein and detecting the presence/absence of a reaction; (f) a method for detecting abnormal stress conditions, including disease, comprising contacting a sample with antibodies specific for **HSF** such that binding occurs; and detection of binding in the nucleus of the cellular sample, etc..

In (b) the vector is pref. a plasmid, bacteriophage or eukaryotic virus vector, esp. plasmid pJC10 or pJC1. The host cell may be prokaryotic, esp. Escherichia coli, or eukaryotic. The antibodies may be mono- or polyclonal. In (e) the reagent is an antibody.

USE/ADVANTAGE - The gene fragments and antibodies may be used to detect abnormal stress conditions in cells. The molecular structure of HSFs and comparison of HSFs between species may be examined.

0/18

L22 ANSWER 21 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1991-252343 [34] WPIDS
DOC. NO. CPI: C1991-109584
TITLE: DNA encoding Drosophila and human heat shock factor proteins - used for developing prods. for studying stress and disease states in living systems.
DERWENT CLASS: B04 D16
INVENTOR(S): CLOS, J; RABINDRAN, S; WESTWOOD, J T; WU, C
PATENT ASSIGNEE(S): (USSH) NAT INST OF HEALTH
COUNTRY COUNT: 1
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
US 7617901	A	19910716	(199134)*		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
US 7617901	A	US 1990-617901	19901126

PRIORITY APPLN. INFO: US 1990-617901 19901126
AB US 7617901 A UPAB: 19930928
The following are disclosed: (A) a DNA segment encoding all or a unique portion of either Drosophila or human heat shock factor (**HSF**)

protein; (B) a recombinantly produced protein encoded by all or a unique portion of the DNA sequences shown; (C) polyclonal and monoclonal antibodies specific for either Drosophila or human **HSF** proteins; (D) a **recombinant** DNA molecule comprising a DNA segment encoding all or a unique portion of either Drosophila or human **HSF** protein and a vector; (E) a host cell stably transformed with the **recombinant** DNA molecule as in (D) in a manner allowing expression of the encoded protein.

USE - The prods. can be used for detection of stress or a diseased state in living systems. They can also be used for identifying HSE genes from other species. The genes can be linked to a tissue-general or tissue-specific promoter and introduced into transgenic mice as a tool for eliciting increased or chronic stress response conditions. Such mice can serve as a biological model for how tissues respond to chronic stress condition e.g. by viral infection, chemical, or mechanical stress. The genes can also be used to increase expression of other gene prods. by contransfecting the **HSF** gene together with other genes linked to **HSF** genes.

0/18

L22 ANSWER 22 OF 24 CAPLUS COPYRIGHT 2000 ACS
ACCESSION NUMBER: 1991:423377 CAPLUS
DOCUMENT NUMBER: 115:23377
TITLE: Three tomato genes code for **heat stress transcription factors** with a region of remarkable homology to the DNA-binding domain of the yeast HSF
AUTHOR(S): Scharf, Klaus Dieter; Rose, Sonja; Zott, Wolfgang; Schoeff, Fritz; Nover, Lutz
CORPORATE SOURCE: Dep. Stress Res., Inst. Plant Biochem., Halle, 4050, Fed. Rep. Ger.
SOURCE: EMBO J. (1990), 9(13), 4495-501
CODEN: EMJODG; ISSN: 0261-4189
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Heat stress (hs) treatment of cell cultures of Lycopersicon peruvianum (Lp, tomato) results in activation of preformed transcription factor(s) (HSF) binding to the heat stress consensus element (HSE). Using appropriate synthetic HSE oligonucleotides, 3 types of clones with potential HSE binding domains were isolated from a tomato .lambda.gt11 expression library by DNA-ligand screening. One of the potential HSF genes is constitutively expressed, the other 3 are hs-induced. Sequence comparison defines a single domain of .apprx.90 amino acid residues common to all 3 genes and to the HSE-binding domain of the yeast HSF. The domain is flanked by proline residues and characterized by 2 long overlapping repeats. The derived consensus sequence may be representative for other eukaryotic HSF and the existence of several different HSF may not be unique to plants.

L22 ANSWER 23 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1989-271986 [38] WPIDS
DOC. NO. NON-CPI: N1989-207768
DOC. NO. CPI: C1989-120384
TITLE: New pure heat shock factor, its activators and DNA encoding sequences - for treatment and diagnosis of diseases and stress associated with heat shock response e.g. hypoxia.
DERWENT CLASS: B04
INVENTOR(S): KINGSTON, R E; SCHUETZ, T J
PATENT ASSIGNEE(S): (GEHO) GEN HOSPITAL CORP
COUNTRY COUNT: 22
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
EP 333201	A	19890920	(198938)*	EN	15
R: AT BE CH DE ES FR GB GR IT LI LU NL SE					
WO 8908661	A	19890921	(198940)	EN	
W: AU DK FI JP KR					
PT 90019	A	19891110	(198950)		
AU 8932804	A	19891005	(199001)		
ZA 8902004	A	19891129	(199002)		
FI 9004564	A	19900917	(199105)		
DK 9002247	A	19901115	(199106)		
JP 03503410	W	19910801	(199137)		
US 5137805	A	19920811	(199235)		13
AU 645243	B	19940113	(199408)		
EP 333201	B1	19940928	(199437)	EN	17
R: AT BE CH DE ES FR GB GR IT LI LU NL SE					
DE 68918479	E	19941103	(199443)		
ES 2064375	T3	19950201	(199511)		
IE 62447	B	19950208	(199518)		

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
EP 333201	A		EP 1989-104763	19890317
WO 8908661	A		WO 1989-US963	19890310
ZA 8902004	A		ZA 1989-2004	19890316
JP 03503410	W		JP 1989-503364	19890310
US 5137805	A	CIP of	US 1988-169965	19880318
			US 1989-301417	19890125
AU 645243	B		AU 1989-32804	19890310
EP 333201	B1		EP 1989-104763	19890317
DE 68918479	E		DE 1989-618479	19890317
			EP 1989-104763	19890317
ES 2064375	T3		EP 1989-104763	19890317
IE 62447	B		IE 1989-817	19890314

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 645243	B	Previous Publ.	AU 8932804
		Based on	WO 8908661
DE 68918479	E	Based on	EP 333201
ES 2064375	T3	Based on	EP 333201

PRIORITY APPLN. INFO: US 1988-169965 19880318; US 1989-301417 19890125

AB EP 333201 A UPAB: 19930923

Heat shock factor (**HSF**), free of natural contaminants, and their derivs. are new. Also new are (1) **recombinant** DNA molecules encoding these cpds. and (2) agents (A) capable of activating **HSF**).

USE - **HSF**, which regulates the expression of hsp genes encoding protective proteins, can be used to treat diseases and stress associated with a heat shock response, e.g. hypoxia or ethanol intoxication. Assay of **HSF** (e.g. by immunoassay or by imaging reaction with labelled antibodies) can be used to diagnose stress and to assess suitability of organs for transplantation. **HSF** can also be used to identify (A) which are themselves useful therapeutically, e.g. in patients with inadequate natural capacity for activating **HSF**.

0/0

L22 ANSWER 24 OF 24 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD
ACCESSION NUMBER: 2000-062154 [05] WPIDS
DOC. NO. CPI: C2000-017188
TITLE: Molecular **circuits** allowing sustained gene expression, useful in protein production and targeted gene therapy.
DERWENT CLASS: B04 D16
INVENTOR(S): VOELLMY, R
PATENT ASSIGNEE(S): (VOEL-I) VOELLMY R
COUNTRY COUNT: 85
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9957290	A1	19991111	(200005)*	EN	51

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ UG ZW
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZA ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9957290	A1	WO 1999-US9748	19990504

PRIORITY APPLN. INFO: US 1998-84236 19980505
AB WO 9957290 A UPAB: 20000128
NOVELTY - New molecular regulatory **circuits** allow sustained expression of a gene of interest using a single application of 'stress' e.g. elevated temperature.

DETAILED DESCRIPTION - Molecular **circuits** comprise:

(i) a polynucleotide comprising a gene encoding a transcription factor and, operably linked to the transcription factor gene, a promoter activatable by stress and a second polynucleotide comprising a gene of interest, the transcription factor gene and, operably linked to the gene of interest and transcription factor gene, a second promoter activatable by the transcription factor;

(ii) as in (i) in which the first promoter in is activatable by stress and by the transcription factor, and the second polynucleotide in comprises a gene of interest and, operably linked to the gene, a second promoter activatable by the transcription factor;

(iii) as in (i) and a second polynucleotide comprising a gene encoding the transcription factor and, operably linked to the gene, a second promoter activatable by the transcription factor and a third polynucleotide comprising a gene of interest and, operably linked to the gene of interest, a third promoter activatable by the transcription factor; and

(iv) as in (iii) in which the gene encodes a second transcription factor and the second promoter is activatable by both first and second transcription factors, and in which the third promoter is activatable by the second transcription factor.

INDEPENDENT CLAIMS are also included for:

(1) an expression vector comprising a molecular **circuit** as above, in which the polynucleotides are comprised in a single nucleic acid molecule in **circuits** (i)-(iii);

(2) sets of two expression vectors comprising first and second

polynucleotides as above respectively, or in which the first comprises the first and second polynucleotides and the second comprises the third polynucleotide;

(3) sets of three expression vectors comprising first, second and third polynucleotides as above respectively;

(4) **insect**, avian, yeast or **mammalian** host cells comprising (1), (2) or (3);

(5) viruses (e.g. adeno-associated viruses) comprising (1);

(6) a method of producing a protein of interest comprising:

(a) culturing the **recombinant** host cells;

(b) stimulating the first promoter by exposing the cultures **recombinant** cells to stress; and

(c) isolating the protein of interest from the cultured **recombinant** host cells where the protein of interest is expressed by the gene of interest;

(7) a method of treating a subject with protein of interest comprising:

(i) administering a pharmaceutical composition to the subject; and

(ii) applying heat to the area of the subject in need of the protein of interest where the heat treatment results in the stimulation of the expression of the gene of interest; and

(8) a method of stimulating the expression of a gene in interest in a **recombinant** cell comprising:

(a) producing a **recombinant** host cell by introducing an expression vector into a host cell; and

(b) exposing the **recombinant** host cell to a condition of stress, where the stress exposure stimulates the first promoter to increase expression of the gene operably linked to the first promoter which in turn results in the stimulation of expression of the gene of interest.

USE - The **circuits** are useful to stimulate sustained expression of a particular gene in **recombinant** cells, by producing the host cells of (4) and exposing them to a stress to stimulate the first promoter. They may be used to produce proteins (e.g. commercially), by culturing the host (especially **mammalian**) cells of (4), stimulating the first promoter by exposing the cells to stress (e.g. by heating the cells; claimed) and isolating the protein. The expression vectors/sets of expression vectors comprising the **circuits** can be included in pharmaceutical compositions (optionally with a carrier and the virus of (5)), useful to treat subjects with a particular protein, by administering the composition and heating the area requiring the protein (all claimed). Such treatment is useful in gene therapy to target delivery to particular areas, so avoiding side effects of systemic delivery and enabling targeted therapy for e.g. cancer, infectious diseases, rheumatoid arthritis etc.

ADVANTAGE - The **circuits** permit sustained activation of expression of a gene of interest by a single stress application; expression induced by a stress promoter is thus maintained beyond the duration of the stress treatment, previously only possible under extreme stress conditions incompatible with cell survival.

Dwg.0/5

FILE 'HOME' ENTERED AT 16:26:46 ON 07 MAR 2000

```
? show files
File 155:MEDLINE(R) 1966-2000/Apr W3
         (c) format only 2000 Dialog Corporation
File  50:CAB Abstracts 1972-2000/Mar
         (c) 2000 CAB International
File  76:Life Sciences Collection 1982-2000/Dec
         (c) 2000 Cambridge Sci Abs
File 358:Current BioTech Abs 1983-1999/Dec
          (c) 1999 DECHEMA
File   5:Biosis Previews(R) 1969-2000/Mar W1
         (c) 2000 BIOSIS
File  73:EMBASE 1974-2000/Feb W3
         (c) 2000 Elsevier Science B.V.
? ds

Set     Items   Description
S1         36   HEAT(W)STRESS(W)TRANSCRIPTION(W)FACTOR?
S2       2133   HSF
S3      79616   CHIMER? OR CHIMAER?
S4    2069748   GENE OR GENES
S5     926374   MUTAT? OR MUTANT?
S6     447488   RECOMBINANT?
S7      77376   CIRCUIT?
S8         14   RD S1 (unique items)
S9       1110   S2 AND (S3-S7)
S10       359   S2 (5N) (S3-S7)
S11        13   S2 (5N)S3
S12       263   S2 (5N)S4
S13        68   S2 (5N)S5
S14        46   S2 (5N)S6
S15         0   S2 (5N)S7
S16         0   S2 AND S7
S17       174   S2(1N)(S4-S6)
S18        63   S10 AND VERTEBRATE?
S19        19   S10 AND INSECT?
S20        66   S10 AND MAMMAL?
S21        47   (S18-S20) AND S17
S22    531552   TRANSCRIPTION?
S23        42   S21 AND S22
S24        25   RD (unique items)
S25        88   S1 OR S11 OR S23
S26        43   RD (unique items)
? t s26/7/1-43

 26/7/1     (Item 1 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

10132227   99370832
  Heat-induced degradation of PER and TIM in Drosophila bearing a
conditional allele of the heat shock ****transcription**** factor gene.
  Sidote D; Edery I
  Center for Advanced Biotechnology and Medicine, Rutgers University,
Piscataway, New Jersey, USA.
  Chronobiol Int (UNITED STATES) Jul 1999, 16 (4) p519-25, ISSN
0742-0528 Journal Code: CYT
  Contract/Grant No.: NS34958, NS, NINDS
  Languages: ENGLISH
  Document type: JOURNAL ARTICLE
  Heat pulses elicit dramatic and rapid decreases in the levels of the D.
melanogaster period (per) and timeless (tim) proteins (i.e., PER and TIM).
To investigate the possible role of the heat shock pathway in this
```

response, we used Drosophila bearing a conditional allele of the ****hsf**** ****gene**** (termed hsf4), which encodes the heat shock ****transcription**** factor (HSF). At all times in a daily cycle, heat-induced decreases in the levels of PER and TIM were similar in wild-type and hsf4 mutant flies. The results strongly suggest that the heat shock pathway contributes little, if any, to the response of the Drosophila circadian clock to heat signals.

26/7/2 (Item 2 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09653637 98363629
Expression levels of heat shock factors are not functionally coupled to the rate of expression of heat shock genes.
Victor M; Benecke BJ
Department of Biochemistry, Ruhr-University Bochum, Germany.
Mol Biol Rep (NETHERLANDS) Jul 1998, 25 (3) p135-41, ISSN 0301-4851
Journal Code: NGW
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The expression patterns of two ****mammalian**** heat shock factors (HSFs) were analysed in cell systems known to reflect an altered heat shock response. For being able to discriminate between the two closely related factors HSF 1 and HSF 2, specific cDNA sequences were cloned and used to generate antisense RNAs as hybridization probes. In general, in various cell lines expression of the two heat shock factors was clearly different. These expression patterns of the ****HSF**** ****genes**** were not influenced by retinoic acid-induced differentiation of human NT2 and mouse F9 teratocarcinoma cells. Generally, HSF 2 expression was extremely low, whereas the significantly higher expression of HSF 1 revealed cell specific differences. The highest expression rates of both HSFs were observed in 293 cells. To examine whether these high levels are involved in the constitutive expression of heat shock genes in these cells, we analysed the binding pattern of 293 cell proteins to the heat shock elements (HSEs). As with other cells, HSE-binding activity in 293 cells was only observed after heat shock treatment. This points to an HSE-independent way for high level expression of heat shock genes in these cells.

26/7/3 (Item 3 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09346659 98062994
Conservation of a stress response: human heat shock ****transcription**** factors functionally substitute for yeast HSF.
Liu XD; Liu PC; Santoro N; Thiele DJ
Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor 48109-0606, USA.
EMBO J (ENGLAND) Nov 3 1997, 16 (21) p6466-77, ISSN 0261-4189
Journal Code: EMB
Contract/Grant No.: CA44059, CA, NCI
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Heat shock factors (HSF) are important eukaryotic stress responsive ****transcription**** factors which are highly structurally conserved from yeast to ****mammals**** . HSFs bind as homotrimers to conserved promoter DNA recognition sites called HSEs. The baker's yeast Saccharomyces cerevisiae possesses a single essential ****HSF**** ****gene****, while distinct ****HSF**** isoforms have been identified in humans. To ascertain

the degree of functional similarity between the yeast and human HSF proteins, human HSF1 and HSF2 were expressed in yeast cells lacking the endogenous ****HSF**** ****gene**** . We demonstrate that human HSF2, but not HSF1, homotrimerizes and functionally complements the viability defect associated with a deletion of the yeast ****HSF**** ****gene****. However, derivatives of hHSF1 that give rise to a trimerized protein, through disruption of a carboxyl- or aminoterminal coiled-coil domain thought to engage in intramolecular interactions that maintain the protein in a monomeric state, functionally substitute for yeast HSF. Surprisingly, hHSF2 expressed in yeast activates target gene ****transcription**** in response to thermal stress. Moreover, hHSF1 and hHSF2 exhibit selectivity for ****transcriptional**** activation of two distinct yeast heat shock responsive genes, which correlate with previously established ****mammalian**** HSF DNA binding preferences in vitro. These results provide new insight into the function of human HSF isoforms, and demonstrate the remarkable functional conservation between yeast and human HSFs, critical ****transcription**** factors required for responses to physiological, pharmacological and environmental stresses.

26/7/4 (Item 4 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09328471 98035041
The GCN4 leucine zipper can functionally substitute for the heat shock transcription factor's trimerization domain.
Drees BL; Grotkopp EK; Nelson HC
University of California, Department of Molecular and Cell Biology, Berkeley 94720-3206, USA.
J Mol Biol (ENGLAND) Oct 17 1997, 273 (1) p61-74, ISSN 0022-2836
Journal Code: J6V
Contract/Grant No.: GM44086, GM, NIGMS; GM08295, GM, NIGMS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The heat shock transcription factor (HSF) is the only known sequence-specific, homotrimeric DNA-binding protein. HSF binds to a DNA recognition site called a heat shock element (HSE), which contains varying numbers of nGAAn units ("GAA boxes") arranged in inverted repeats. To investigate the role of trimerization on HSF's DNA-binding properties, we replaced the trimerization domain, which self-assembles to form a three-stranded alpha-helical coiled coil, with the GCN4 leucine zipper, which forms a two-stranded alpha-helical coiled coil. Surprisingly, this substitution did not effect the ability of HSF to function in vivo. Biochemical studies of an ****HSF****-leucine zipper ****chimera**** in comparison to an ****HSF**** truncation show that the ****HSF****-leucine zipper ****chimera**** , though dimeric in solution and dimeric when bound to a two-box HSE, forms a trimeric complex when bound to a three-box HSE. The ability to form trimers depends on the presence of three contiguous GAA boxes present in inverted repeats. The proximity of the leucine zippers due to the orientation of the binding sites suggests that the leucine zippers might be forming a three-stranded coiled coil and several experiments lend support to this model. The ability of the leucine zipper to change oligomeric states in context might explain why the leucine zipper can replace the trimerization domain of HSF in vivo.

26/7/5 (Item 5 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09182002 97411910

****Heat**** ****stress**** ****transcription**** ****factors**** from tomato can functionally replace HSF1 in the yeast Saccharomyces cerevisiae.
Boscheinen O; Lyck R; Queitsch C; Treuter E; Zimarino V; Scharf KD
Molecular Cell Biology, Biocenter of the J.W. Goethe University, Frankfurt, Germany.
Mol Gen Genet (GERMANY) Jul 1997, 255 (3) p322-31, ISSN 0026-8925 Journal Code: NGP
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The fact that yeast HSF1 is essential for survival under nonstress conditions can be used to test heterologous Hsfs for the ability to substitute for the endogenous protein. Our results demonstrate that like Hsf of Drosophila, tomato Hsfs A1 and A2 can functionally replace the corresponding yeast protein, but Hsf B1 cannot. In addition to survival at 28 degrees C, we checked the transformed yeast strains for temperature sensitivity of growth, induced thermotolerance and activator function using two different lacZ reporter constructs. Tests with full-length Hsfs were supplemented by assays using mutant Hsfs lacking parts of their C-terminal activator region or oligomerization domain, or containing amino acid substitutions in the DNA-binding domain. Remarkably, results with the yeast system are basically similar to those obtained by the analysis of the same Hsfs as transcriptional activators in a tobacco protoplast assay. Most surprising is the failure of HsfB1 to substitute for the yeast Hsf. The defect can be overcome by addition to HsfB1 of a short C-terminal peptide motif from HsfA2 (34 amino acid residues), which represents a type of minimal activator necessary for interaction with the yeast transcription apparatus. Deletion of the oligomerization domain (HR-A/B) does not interfere with Hsf function for survival or growth at higher temperatures. But monomeric Hsf has a markedly reduced affinity for DNA, as shown by lacZ reporter and band-shift assays.

26/7/6 (Item 6 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09163935 97365780
The Hsf world: classification and properties of plant ****heat**** ****stress**** ****transcription**** ****factors****.
Nover L; Scharf KD; Gagliardi D; Vergne P; Czarnecka-Verner E; Gurley WB
Biocenter of the Goethe University, Frankfurt/M., Germany. Nover@cellbiology.uni-frankfurt.d400.de
Cell Stress Chaperones (UNITED STATES) Dec 1996, 1 (4) p215-23, ISSN 1355-8145 Journal Code: CV5
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, TUTORIAL
Based on the partial or complete sequences of 14 plant ****heat**** ****stress**** ****transcription**** ****factors**** (Hsfs) from tomato, soybean, Arabidopsis and maize we propose a general nomenclature with two basic classes, i.e. classes A and B each containing two or more types of Hsfs (HsfA1, HsfA2 etc.). Despite some plant-specific peculiarities, essential functional domains and modules of these proteins are conserved among plants, yeast, Drosophila and ****vertebrates**** . A revised terminology of these parts follows recommendations agreed upon among the authors and representatives from other laboratories working in this field (see legend to Fig. 1). Similar to the situation with the small heat shock proteins (sHsps), the complexity of the ****hsf**** ****gene**** family in plants appears to be higher than in other eukaryotic organisms. (58 Refs.)

26/7/7 (Item 7 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

(c) format only 2000 Dialog Corporation. All rts. reserv.

09155378 97320172
Intracellular distribution and identification of the nuclear localization signals of two plant ****heat****-****stress**** ****transcription**** ****factors****.
Lyck R; Harmening U; Hohfeld I; Treuter E; Scharf KD; Nover L
Department of Molecular and Cellular Biology, Biocenter J.W. Goethe-University, Frankfrut/Main, Germany.
Planta (GERMANY) 1997, 202 (1) p117-25, ISSN 0032-0935 Journal Code: BNG
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Similar to ****heat****-****stress**** ****transcription**** ****factors**** (HSFs) from non-plant sources, HSFA1 and HSFA2 from tomato (Lycopersicon esculentum Mill) contain two conserved clusters of basic amino acid residues (K/R1 and K/R2) which might serve as nuclear localization signal (NLS) motifs. Mutation of either one of them and functional testing of the corresponding proteins in a transient expression assay using tobacco (Nicotiana plumbaginifolia L:) protoplasts gave the following results. Whereas K/R1, positioned in all HSFs at the C-terminus of the DNA-binding domain, had no influence on nuclear import, the K/R1 mutants were impaired in their interaction with the DNA (band-shift assays). In contrast to this, mutants of the K/R2 motif, found 15-20 amino acid residues C-terminal of the oligomerization domain (HR-A/B region), had wild-type activity in DNA-binding but were defective in nuclear import. Thus, for the related tomato HSFA1 and HSFA2 the K/R2 cluster represents the only NLS motif, and in this function it cannot be replaced by K/R1.

26/7/8 (Item 8 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09111814 97326118
Different thresholds in the responses of two heat shock ****transcription**** factors, HSF1 and HSF3.
Tanabe M; Nakai A; Kawazoe Y; Nagata K
Department of Cell Biology, Chest Disease Research Institute, Kyoto University, Sakyo-Ku, Kyoto 606-01, Japan.
J Biol Chem (UNITED STATES) Jun 13 1997, 272 (24) p15389-95, ISSN 0021-9258 Journal Code: HIV
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Avian cells express three ****HSF**** ****genes**** encoding a unique factor, HSF3, as well as homologues of ****mammalian**** HSF1 and HSF2. HSF1 is the major factor that mediates the heat shock signal in ****mammalian**** cells. We reported previously that cHSF3, as well as cHSF1, is activated by heat shock in chicken cells. In this study, we examined the functional differences between cHSF1 and cHSF3. Comparison of the heat-inducible DNA binding activity of cHSF1 with cHSF3 at various temperatures revealed that the latter was activated at higher temperatures than the former. At a mild heat shock, such as 41 degrees C, only cHSF1 was activated, whereas both cHSF1 and cHSF3 were activated following a severe heat shock at 45 degrees C. Heat-inducible nuclear translocation and trimerization were accompanied by DNA binding activity. We also observed that cHSF3 was activated by treating cells with higher concentrations of sodium arsenite compared to cHSF1. The DNA binding activity of cHSF3 by severe heat shock lasted for a longer period than that of cHSF1. Interestingly, the total amount of cHSF3 increased only upon severe heat shock, whereas that of HSF1 decreased. Substantial amounts of cHSF3 remained in the soluble fraction under severe heat shock, whereas cHSF1

rapidly moved to the insoluble fractions in that conditions. Comparison of ****transcriptional**** activity of the activation domains of cHSF1 and cHSF3 revealed that the activity of cHSF3 was as strong as that of cHSF1. These findings indicate that there are different thresholds for cHSF1 and cHSF3 and that cHSF3 is involved in the persistent and burst activation of stress genes upon severe stress in chicken cells. Pretreatment of cycloheximide elevated the threshold concentrations of arsenite of both factors. This suggests that denaturation of nascent polypeptides could be the first trigger for the activation of both factors, and the pathways for activation of cHSF1 and cHSF3 may be identical, or at least share some common mechanisms.

26/7/9 (Item 9 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

08903834 97127404
HSF4, a new member of the human heat shock factor family which lacks properties of a ****transcriptional**** activator.
Nakai A; Tanabe M; Kawazoe Y; Inazawa J; Morimoto RI; Nagata K
Department of Cell Biology, Kyoto University, Japan. nakai@chest.kyoto-u.ac.jp
Mol Cell Biol (UNITED STATES) Jan 1997, 17 (1) p469-81, ISSN 0270-7306 Journal Code: NGY
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Heat shock ****transcription**** factors (HSFs) mediate the inducible ****transcriptional**** response of genes that encode heat shock proteins and molecular chaperones. In ****vertebrates****, three related ****HSF**** ****genes**** (HSF1 to -3) and the respective gene products (HSFs) have been characterized. We report the cloning and characterization of human HSF4 (hHSF4), a novel member of the hHSF family that shares properties with other members of the HSF family yet appears to be functionally distinct. hHSF4 lacks the carboxyl-terminal hydrophobic repeat which is shared among all ****vertebrate**** HSFs and has been suggested to be involved in the negative regulation of DNA binding activity. hHSF4 is preferentially expressed in the human heart, brain, skeletal muscle, and pancreas. Transient transfection of hHSF4 in HeLa cells, which do not express hHSF4, results in a constitutively active DNA binding trimer which, unlike other members of the HSF family, lacks the properties of a ****transcriptional**** activator. Constitutive overexpression of hHSF4 in HeLa cells results in reduced expression of the endogenous hsp70, hsp90, and hsp27 genes. hHSF4 represents a novel hHSF that exhibits tissue-specific expression and functions to repress the expression of genes encoding heat shock proteins and molecular chaperones.

26/7/10 (Item 10 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

08827131 97009869
The ****transcriptional**** regulation of heat shock genes: a plethora of heat shock factors and regulatory conditions.
Morimoto RI; Kroeger PE; Cotto JJ
Department of Biochemistry, Molecular Biology and Cell Biology Northwestern University, Evanston, IL 60208, USA.
EXS (SWITZERLAND) 1996, 77 p139-63, Journal Code: BFZ
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, TUTORIAL
The inducible regulation of heat shock gene ****transcription**** is

mediated by a family of heat shock factors (HSF) that respond to diverse forms of physiological and environmental stress including elevated temperature, amino acid analogs, heavy metals, oxidative stress, anti-inflammatory drugs, arachidonic acid, and a number of pathophysiological disease states. The ****vertebrate**** genome encodes a family of HSFs which are expressed ubiquitously, yet the DNA binding properties of each factor are negatively regulated and activated in response to specific conditions. This chapter will discuss the regulation of the ****HSF**** multi-****gene**** family and the role of these ****transcriptional**** activators in the inducible expression of genes encoding heat shock proteins and molecular chaperones. (116 Refs.)

26/7/11 (Item 11 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

08709022 96270744
Solution structure of the DNA-binding domain of the tomato ****heat****-****stress**** ****transcription**** ****factor**** HSF24.
Schultheiss J; Kunert O; Gase U; Scharf KD; Nover L; Ruterjans H
Department of Biophysical Chemistry, Biocenter of the Goethe-Universitat, Frankfurt, Germany.
Eur J Biochem (GERMANY) Mar 15 1996, 236 (3) p911-21, ISSN 0014-2956
Journal Code: EMZ
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Two-dimensional-NMR and three-dimensional-NMR experiments were performed to determine the solution structure of the DNA-binding domain of the tomato ****heat****-****stress**** ****transcription**** ****factor**** HSF24. Samples of uniformly 15N-labeled and 15N, 13C-labeled recombinant proteins were used in the investigation. A near-complete assignment of the backbone 1H, 15N, and 13C resonances was obtained by three-dimensional triple-resonance experiments, whereas three-dimensional 15N-TOCSY-heteronuc lear-single-quantum-correlation-spectroscopy, HCCH-COSY and HCCH-TOCSY spectra were recorded for side-chain assignments, 885 non-redundant distance constraints from two-dimensional-homonuclear and three-dimensional-15N-edited and 13C-edited NOESY spectra and 40 hydrogen-bond constraints from exchange experiments were used for structure calculations. The resulting three-dimensional structure contains a three-helix bundle and a small four-stranded antiparallel beta-sheet that forms a hydrophobic core. The two C-terminal helices are parts of a highly conserved helix-turn-helix motif that is probably involved in DNA recognition and binding. In contrast to heat-stress factors from yeast and animals, the plant heat-stress factors lack a loop of 11 amino acid residues inserted between beta3 and beta4. This leads to a tight turn between these beta-strands.

26/7/12 (Item 12 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

08484801 96039624
Derepression of the activity of genetically engineered heat shock factor causes constitutive synthesis of heat shock proteins and increased thermotolerance in transgenic Arabidopsis.
Lee JH; Hubel A; Schoffl F
Lehrstuhl fur Allgemeine Genetik, Universitat Tubingen, Germany.
Plant J (ENGLAND) Oct 1995, 8 (4) p603-12, ISSN 0960-7412
Journal Code: BRU
Languages: ENGLISH

Document type: JOURNAL ARTICLE
ATHSF1 is a heat shock transcription factor (HSF) of Arabidopsis that is constitutively expressed but its activity for DNA binding, trimer formation and transcriptional activation of heat shock (hs) genes is repressed at normal temperatures. In this study the functional properties of ****chimeric**** ****HSF**** -glucuronidase (GUS) fusion proteins were tested. Ectopic expression of HSF-GUS or GUS-HSF in transgenic Arabidopsis plants resulted in a derepression of HSF functions as shown by trimer formation, specific DNA binding, and the constitutive expression of heat shock proteins (HSPs) at normal temperature. A novel GUS activity-staining protocol was used to show the specific binding of trimeric HSF fusion proteins to DNA and following hs, an interaction between ****chimeric**** ****HSF****-GUS and authentic ****HSF**** proteins. The ****chimeric**** HSFs were insensitive to the negative regulation that counteracts activation of the authentic HSF at normal temperature. Heterotrimer complexes were reconstituted in vitro from recombinant ATHSF1 and HSF-GUS proteins expressed in Escherichia coli and using this protocol, the temperature-dependent activation of wt HSF was monitored in vivo and in vitro. Transgenic plants expressing constitutively active HSF-GUS fusion proteins are also constitutive for HSPs. Approximately 20% of the maximum heat-inducible levels of HSP18 were already present at normal temperature. The level of basic thermotolerance was significantly enhanced in these plants. The results indicate that genetic engineering using protein fusion is a very effective means to derepress the activity of an important regulatory protein in plants, that consequently activates a constitutive hs response in the absence of heat stress and eventually alters the thermotolerance phenotype.

26/7/13 (Item 13 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

07716442 94105354
Two cDNAs for tomato ****heat**** ****stress**** ****transcription**** ****factors****.
Scharf KD; Rose S; Thierfelder J; Nover L
Institute of Plant Biochemistry, Halle, Germany.
Plant Physiol (UNITED STATES) Aug 1993, 102 (4) p1355-6, ISSN 0032-0889 Journal Code: P98
Languages: ENGLISH
Document type: JOURNAL ARTICLE

26/7/14 (Item 14 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

07596742 93341449
Promoter specificity and deletion analysis of three ****heat**** ****stress**** ****transcription**** ****factors**** of tomato.
Treuter E; Nover L; Ohme K; Scharf KD
Institute of Plant Biochemistry, Halle, Germany.
Mol Gen Genet (GERMANY) Jul 1993, 240 (1) p113-25, ISSN 0026-8925 Journal Code: NGP
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Transient expression assays in transformed tobacco (Nicotiana plumbaginifolia) mesophyll protoplasts were used to test the activity of three tomato ****heat**** ****stress**** ****transcription**** ****factors**** , HSF24, HSF8 and HSF30, in a trans-activation and a trans-repression assay. The results document differences between the three

HSFs with respect to their response to the configuration of heat stress promoter elements (HSEs) in the reporter construct (promoter specificity) and to the stress regime used for activation. Analysis of C-terminal deletions identified acidic sequence elements with a central tryptophan residue, which are important for HSF activity control. Surprisingly, heterologous HSFs from Drosophila and human cells, but not from yeast, were also functional as heat stress-induced transcription factors in this tobacco protoplast system.

26/7/15 (Item 15 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

07517942 93204945
Characterization of a novel chicken heat shock ****transcription**** factor, heat shock factor 3, suggests a new regulatory pathway.
Nakai A; Morimoto RI
Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208.
Mol Cell Biol (UNITED STATES) Apr 1993, 13 (4) p1983-97, ISSN 0270-7306 Journal Code: NGY
Contract/Grant No.: GM38109, GM, NIGMS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
We have cloned three avian heat shock ****transcription**** factor (****HSF****) ****genes**** corresponding to a novel factor, HSF3, and the avian homologs of ****mammalian**** HSF1 and HSF2. The predicted amino acid sequence of HSF3 is approximately 40% related to the sequence of HSF1 and HSF2. The sequences for all three factors exhibit extensive identify in the DNA binding motifs and the heptad repeats of hydrophobic amino acids which are common to all eukaryotic HSFs. Despite these overall similarities, each avian HSF exhibits distinct DNA binding properties. HSF2 when expressed in vitro binds constitutively to the heat shock element promoter sequence, whereas neither HSF1 nor HSF3 expressed in vitro binds to DNA. HSF1 DNA binding is induced upon heat shock or treatment with nonionic detergents, whereas the DNA binding properties of HSF3 are not induced by these conditions in vitro. These results suggest that HSF3 activation may involve an induction pathway distinct from the traditional forms of heat shock gene induction. HSF3 DNA binding activity, however, is obtained when the carboxyl-terminal region including the distal heptad repeat is deleted, indicating the presence of negative cis-regulatory sequences. The HSF3 message, like HSF1 and HSF2 messages, is coexpressed during development and in most tissues, which suggests a general role for the regulatory pathway involving HSF3.

26/7/16 (Item 16 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

06949882 92009187
Vectors for the expression and analysis of DNA-binding proteins in yeast.
Bonner JJ
Department of Biology, Indiana University, Bloomington 47405.
Gene (NETHERLANDS) Jul 31 1991, 104 (1) p113-8, ISSN 0378-1119 Journal Code: FOP
Contract/Grant No.: GM26693, GM, NIGMS; RR7031-25, RR, NCRR
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A series of 13 vectors is described. All are yeast centromere plasmids with the LEU2 gene for selection in yeast, and pUC19 sequences for growth

in Escherichia coli. All contain the GAL1 promoter directing transcription into a multiple cloning site (MCS). For twelve of the plasmids, synthetic oligodeoxyribonucleotides create an ATG start codon, in a productive context for yeast, prior to the MCS. Spacing between the ATG and the MCS is variable, to facilitate the cloning of gene fragments in the appropriate reading frame. Nine of the plasmids also contain the strong transcriptional activator from the herpes simplex virus VP16 gene, joined downstream from the MCS. In these nine vectors, all possible combinations of reading frames are available. The suitability of these plasmids for the expression and analysis of DNA-binding domains is tested by cloning into them fragments of the yeast HSF1 gene, encoding the heat shock transcription factor (HSF). The regulation of reporter gene expression by the ****chimeric**** ****HSF**** -VP16 fusions is described, as is the utility of these vectors for other applications.

26/7/17 (Item 17 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

06625395 91092274
Three tomato genes code for ****heat**** ****stress**** ****transcription**** ****factors**** with a region of remarkable homology to the DNA-binding domain of the yeast HSF [published erratum appears in EMBO J 1991 Apr;10(4):1026]
Scharf KD; Rose S; Zott W; Schoffl F; Nover L; Schoff F [corrected to Schoffl F]
Department of Stress Research, Institute of Plant Biochemistry, Halle, FRG.
EMBO J (ENGLAND) Dec 1990, 9 (13) p4495-501, ISSN 0261-4189 Journal Code: EMB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Heat stress (hs) treatment of cell cultures of Lycopersicon peruvianum (Lp, tomato) results in activation of preformed transcription factor(s) (HSF) binding to the heat stress consensus element (HSE). Using appropriate synthetic HSE oligonucleotides, three types of clones with potential HSE binding domains were isolated from a tomato lambda gt11 expression library by DNA-ligand screening. One of the potential HSF genes is constitutively expressed, the other two are hs-induced. Sequence comparison defines a single domain of approximately 90 amino acid residues common to all three genes and to the HSE--binding domain of the yeast HSF. The domain is flanked by proline residues and characterized by two long overlapping repeats. We speculate that the derived consensus sequence is also representative for other eukaryotic HSF and that the existence of several different HSF is not unique to plants.

26/7/18 (Item 1 from file: 50)
DIALOG(R)File 50:CAB Abstracts
(c) 2000 CAB International. All rts. reserv.

03093089 CAB Accession Number: 951610772
Arabidopsis heat shock factor is constitutively active in Drosophila and human cells.
Hubel, A.; Lee, J. H.; Wu, C.; Schoffl, F.
Universitat Tubingen, Biologisches Institut, Lehrstuhl fur Allgemeine Genetik, Auf der Morgenstelle 28, D-72076 Tubingen, Germany.
Molecular and General Genetics vol. 248 (2): p.136-141
Publication Year: 1995
ISSN: 0026-8925
Language: English

Document Type: Journal article
Heat shock factors (HSF) are the ****transcriptional**** activators of the heat shock response. The conversion of constitutively expressed HSF to a form that can bind DNA requires the trimerization of the protein, involving leucine zipper interactions, as shown for yeast, Drosophila, chicken and human HSFs. Like other metazoan HSFs, the endogenous Arabidopsis HSF displays heat shock-inducible DNA-binding activity in gel retardation assays. The heat shock-inducible binding of a ****recombinant**** Arabidopsis ****HSF**** (ATHSF1) expressed in Arabidopsis plants suggests that ATHSF1 is the major HSF regulating the heat stress response. However, on transient expression in Drosophila and human cells, ATHSF1 fails to exhibit proper regulation, as demonstrated by constitutive binding to DNA, and by constitutive expression of a chloramphenicol acetyltransferase (CAT) reporter gene under the control of the Drosophila hsp70 promoter. These results suggested that the regulation of ATHSF1 is normally dependent on a specific factor that inhibits the DNA-binding and ****transcriptional**** activities under non-heat shock conditions. 28 ref.

26/7/19 (Item 2 from file: 50)
DIALOG(R)File 50:CAB Abstracts
(c) 2000 CAB International. All rts. reserv.

03010278 CAB Accession Number: 951604940
Heat stress promoters and transcription factors.
Scharf, K. D.; Materna, T.; Treuter, E.; Nover, L.
Lehrstuhl Zellbiologie, Biozentrum, J.-W.-Goethe-Universitat, 60439 Frankfurt, Germany.
Book Title: Plant promoters and transcription factors.
p.125-162
Publication Year: 1994
Results and Problems in Cell Differentiation Volume 20.
Editors: Nover, L.
Publisher: Springer-Verlag GmbH & Co. KG Berlin, Germany
ISBN: 3-540-57288-0
Language: English
Document Type: Book chapter
The subject is examined under these headings: heat stress response and stress protein families; heat stress promoters; cloning of ****heat**** ****stress**** ****transcription**** ****factor**** (hsf) genes; characteristics of transcription factor clones and proteins; the DNA-binding domain; leucine zipper-type hydrophobic repeats; the C-terminal activation domain; stress-induced expression of tomato HSFs; functional analysis of hsf clones in tobacco protoplasts; trans-activation vs. trans-repression assays; deletion analysis of hsf clones; survey of the HSF world 1993; multiplicity and selectivity of ****heat**** ****stress**** ****transcription**** ****factors****; HSF activation and the role of the oligomerization state; and the missing link(s) - model of control of HSF activity. 6 pp. of ref.

26/7/20 (Item 3 from file: 50)
DIALOG(R)File 50:CAB Abstracts
(c) 2000 CAB International. All rts. reserv.

02724122 CAB Accession Number: 931639908
The heat shock response in transgenic plants: the use of chimaeric heat shock genes.
Schoffl, F.; Diedring, V.; Kliem, M.; Rieping, M.; Schroder, G.; Severin, K.
Department of Genetics, University of Tubingen, Auf der Morgenstelle 28,

7400 Tubingen, Germany.
Book Title: Inducible plant proteins: their biochemistry and molecular biology
p.247-266
Publication Year: 1992
Society for Experimental Biology Seminar Series 49
Editors: Wray, J. L.
Publisher: Cambridge University Press Cambridge, UK
ISBN: 0-521-40170-4
Language: English
Document Type: Book chapter
Manipulation of the heat shock (hs) response in plants by genetic engineering will contribute significantly to an understanding of the molecular mechanisms underlying stress-related control of gene expression. Recent advances in this subject are discussed with reference to research on soyabean hs genes under the following headings: (1) functional analysis of hs promoter elements; (2) cis-active scaffold attachment region (SAR) sequences and hs gene expression; (3) constitutive expression of hs genes; (4) generation of antisense hsRNA; (5) hs transcription factor, ****HSF****; and (6) ****HSF****-dependent ****chimaeric**** hs genes as selection markers. It is concluded that: certain hs proteins (hsp) are developmentally regulated and translational control appears to dominate hsp expression in certain developmental stages; proximity of SAR sequences to hs genes implies a correlation between scaffold attachment and gene expression; and the most conserved regions of HSF primary nucleotide and amino acid sequences are the DNA-binding domain (approx equal to 50% identity exists between yeast, Drosophila, mammals and plants) and an array of heptad repeats of hydrophobic residues representing a leucine zipper motif. 42 ref.

26/7/21 (Item 1 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

12036779 BIOSIS NO.: 199900317298
Modulation of human heat shock factor trimerization by the linker domain.
AUTHOR: Liu Phillip C C; Thiele Dennis J(a)
AUTHOR ADDRESS: (a)Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI, 48109-**USA
JOURNAL: Journal of Biological Chemistry 274 (24):p17219-17225 June 11, 1999
ISSN: 0021-9258
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English
SUMMARY LANGUAGE: English

ABSTRACT: Heat shock ****transcription**** factors (HSFs) are stress-responsive proteins that activate the expression of heat shock genes and are highly conserved from bakers' yeast to humans. Under basal conditions, the human HSF1 protein is maintained as an inactive monomer through intramolecular interactions between two coiled-coil domains and interactions with heat shock proteins; upon environmental, pharmacological, or physiological stress, HSF1 is converted to a homotrimer that binds to its cognate DNA binding site with high affinity. To dissect regions of HSF1 that make important contributions to the stability of the monomer under unstressed conditions, we have used functional complementation in bakers' yeast as a facile assay system. Whereas wild-type human HSF1 is restrained as an inactive monomer in yeast that is unable to substitute for the essential yeast ****HSF**** protein, ****mutations**** in the linker region between the DNA binding

domain and the first coiled-coil allow HSF1 to homotrimerize and rescue the viability defect of a hsfDELTA strain. Fine mapping by functional analysis of HSF1-HSF2 chimeras and point mutagenesis revealed that a small region in the amino-terminal portion of the HSF1 linker is required for maintenance of HSF1 in the monomeric state in both yeast and in transfected human 293 cells. Although linker regions in ****transcription**** factors are known to modulate DNA binding specificity, our studies suggest that the human HSF1 linker plays no role in determining HSF1 binding preferences in vivo but is a critical determinant in regulating the HSF1 monomer-trimer equilibrium.

26/7/22 (Item 2 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

11941540 BIOSIS NO.: 199900187649
Regulatory domain of human heat shock ****transcription**** factor-2 is not regulated by hemin or heat shock.
AUTHOR: Zhu Zhen; Mivechi Nahid F(a)
AUTHOR ADDRESS: (a)Medical College of Georgia, Institute of Molecular Medicine and Genetics, 1120 15th St. CB2803, **USA
JOURNAL: Journal of Cellular Biochemistry 73 (1):p56-69 April 1, 1999
ISSN: 0730-2312
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Heat shock ****transcription**** factor 2 (HSF-2) activates ****transcription**** of heat shock proteins in response to hemin in the human erythroleukemia cell line, K562. To understand the regulation of HSF-2 activation, a series of deletion ****mutants**** of ****HSF****-2 fused to the GAL-4 DNA binding domain were generated. We have found that human HSF-2 has a regulatory domain located in the carboxyl-terminal portion of the protein which represses the activity of its activation domain under normal physiological conditions. The repressive effects of this domain can be eliminated by its deletion in GAL4-HSF-2 fusion constructs. The regulatory domain of HSF-2 can also repress a heterologous chimeric activator that contains a portion of the VP16 activation domain. The activation domain of HSF-2 is a segment of approximately 77 amino acids located proximal to the carboxyl-terminal hydrophobic heptad repeat (leucine zipper 4) of the molecule. Interestingly, the GAL4-HSF-2 fusion protein and the 77 amino acids activation domain are inactive and are not activated by pretreatment of cells with either hemin or elevated temperature. Our data suggest that regulation of HSF-2 differs from HSF-1 in that its regulatory domain is not responsive to hemin or heat directly.

26/7/23 (Item 3 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

11850246 BIOSIS NO.: 199900096355
Heat stress response and ****heat**** ****stress**** ****transcription**** ****factors****.
AUTHOR: Scharf Klaus-Dieter; Hoehfeld Ingo; Nover Lutz(a)
AUTHOR ADDRESS: (a)Mol. Cell Biol., Biocent., Goethe-Univ-Frankfurt, Marie-Curie-Str. 9, D-60439 Frankfurt/Main**Germany
JOURNAL: Journal of Biosciences (Bangalore) 23 (4):p313-329 Oct., 1998
ISSN: 0250-5991
DOCUMENT TYPE: Literature Review

RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Expression of heat shock protein (HSP)-coding genes is controlled by ****heat**** ****stress**** ****transcription**** ****factors**** (Hsfs). They are structurally and functionally conserved throughout the eukaryotic kingdom. In addition to the DNA-binding domain with the helix-turn-helix motif essential for DNA recognition, three functional parts in the C-terminal activator domain were characterized: (i) the HR-A/B region is responsible for oligomerization and activity control, (ii) the nuclear localizing signal (NLS) formed by a cluster of basic amino acid residues which is required and sufficient for nuclear import and (iii) short C-terminal peptide motifs with a central Trp residue (AHA elements). These three parts are indispensible for the activator function. A peculiarity of plants is the heat shock-inducible new synthesis of Hsfs. In tomato HsfA1 is constitutively expressed, whereas Hsfs A2 and B1 are heat shock-inducible proteins themselves. We used Hsf knock-out strains of yeast and transient reporter assays in tobacco protoplasts for functional analysis of Hsf-coding cDNA clones and mutants derived from them. HsfA2, which in tomato cell cultures is expressed only after heat shock induction, tends to form large cytoplasmic aggregates together with other HSPs (heat stress granules). In the transient expression assay its relatively low activator potential is evidently due to the inefficient nuclear import. However, the intramolecular shielding of the NLS can be released either by deletion of a short C-terminal fragment or by coexpression with HsfA1, which forms hetero-oligomers with HsfA2.

26/7/24 (Item 4 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

11782231 BIOSIS NO.: 199900028340
Structural organization and promoter analysis of murine heat shock ****transcription**** factor-1 gene.
AUTHOR: Zhang Yan; Koushik Srinagesh; Dai Rujuan; Mivechi Nahid F(a)
AUTHOR ADDRESS: (a)Medical Coll. Georgia, Inst. Molecular Med. Genetics, Gene Regulation Group, 1120 15th St., CB28**USA
JOURNAL: Journal of Biological Chemistry 273 (49):p32514-32521 Dec. 4, 1998
ISSN: 0021-9258
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Heat shock factor-1 (HSF-1) activates ****transcription**** of heat shock proteins in eukaryotes. Several overlapping genomic clones containing the murine ****HSF****-1 ****gene**** were isolated from a phage genomic library. Results indicate that the ****HSF****-1 ****gene**** contains 13 exons that span at least 30 kilobase pairs. Sequence analysis of the 5'-untranslated region of HSF-1 suggests that it contains sequences of a recently described Bop1 gene in reverse orientation within its first 331 base pairs (bp) upstream of the translation initiation site. The minimal promoter sequence required for HSF-1 basal expression was identified by deletion analysis from -4 kilobase pairs to -331 bp of the promoter fused to a luciferase reporter gene using transient transfection assays. Results indicate that 331 hp upstream of the HSF-1 translation start site is required for maximal basal expression in NEH3T3 and F9 cells. This fragment also results in high levels of luciferase activity in the reverse orientation, that is, 5' to the Bop1 gene, suggesting that this segment is bidirectional and

could be utilized for basal expression of both ****HSF****-1 and Bop1 ****genes****. This segment of the promoter contains recognition elements for Sp1 and CCAAT-box binding ****transcription**** factors, which when mutated in either sense or antisense orientations to the ****HSF****-1 ****gene**** results in a reduction of basal expression by 50-75% relative to wild type, suggesting that these sites are critical for basal expression of both ****HSF****-1 and Bop1 ****genes****.

26/7/25 (Item 5 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

11426822 BIOSIS NO.: 199800208154
The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules.
AUTHOR: Scharf Klaus-Dieter; Heider Harald; Hoehfeld Ingo; Lyck Ruth; Schmidt Enrico; Nover Lutz(a)
AUTHOR ADDRESS: (a)Dep. Mol. Cell Biol., Biocenter N200, 30G, Goethe Univ. Frankfurt, Marie Curie Str. 9, D-60439 F**Germany
JOURNAL: Molecular and Cellular Biology 18 (4):p2240-2251 April, 1998
ISSN: 0270-7306
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: In heat-stressed (HS) tomato (Lycopersicon peruvianum) cell cultures, the constitutively expressed HS transcription factor HsfA1 is complemented by two HS-inducible forms, HsfA2 and HsfB1. Because of its stability, HsfA2 accumulates to fairly high levels in the course of a prolonged HS and recovery regimen. Using immunofluorescence and cell fractionation experiments, we identified three states of HsfA2: (i) a soluble, cytoplasmic form in preinduced cultures maintained at 25degree C, (ii) a salt-resistant, nuclear form found in HS cells, and (iii) a stored form of HsfA2 in cytoplasmic HS granules. The efficient nuclear transport of HsfA2 evidently requires interaction with HsfA1. When expressed in tobacco protoplasts by use of a transient-expression system, HsfA2 is mainly retained in the cytoplasm unless it is coexpressed with HsfA1. The essential parts for the interaction and nuclear cotransport of the two Hsfs are the homologous oligomerization domain (HR-A/B region of the A-type Hsfs) and functional nuclear localization signal motifs of both partners. Direct physical interaction of the two Hsfs with formation of relatively stabile hetero-oligomers was shown by a two-hybrid test in Saccharomyces cerevisiae as well as by coimmunoprecipitation using tomato and tobacco whole-cell lysates.

26/7/26 (Item 6 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10943983 BIOSIS NO.: 199799565128
Multiple functions of Drosophila heat shock ****transcription**** factor in vivo.
AUTHOR: Judlicka Paul; Mortin Mark A; Wu Carl(a)
AUTHOR ADDRESS: (a)Lab. Mol. Cell Biol., Natl. Cancer Inst., Build. 37, Room 5E-26, Bethesda, MD 20892**USA
JOURNAL: EMBO (European Molecular Biology Organization) Journal 16 (9):p 2452-2462 1997
ISSN: 0261-4189
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Heat shock ****transcription**** factor (HSF) is a ****transcriptional**** activator of heat shock protein (hsp) genes in eukaryotes. In order to elucidate the physiological functions of HSF in Drosophila, we have isolated lethal ****mutations**** in the ****hsf**** ****gene****. Using a conditional allele, we show that HSF has an essential role in the ability of the organism to survive extreme heat stress. In contrast to previous results obtained with yeast HSF, the Drosophila protein is dispensable for general cell growth or viability. However, it is required under normal growth conditions for oogenesis and early larval development. These two developmental functions of Drosophila HSF are genetically separable and appear not to be mediated through the induction of HSPs, implicating a novel action of HSF that may be unrelated to its characteristic function as a stress-responsive ****transcriptional**** activator.

26/7/27 (Item 7 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10693716 BIOSIS NO.: 199799314861
Regulation of Drosophila heat shock factor trimerization: Global sequence requirements and independence of nuclear localization.
AUTHOR: Orosz Andras; Wisniewski Jan; Wu Carl(a)
AUTHOR ADDRESS: (a)Lab. Molecular Cell Biol., Natl. Cancer Inst., Build. 37, Room 4C-09, Bethesda, MD 20892-4255**USA
JOURNAL: Molecular and Cellular Biology 16 (12):p7018-7030 1996
ISSN: 0270-7306
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Heat shock ****transcription**** factor (HSF) is a multidomain protein that exists as a monomer under normal conditions and is reversibly induced upon heat shock to a trimeric state that binds to DNA with high affinity. The maintenance of the monomeric state is dependent on hydrophobic heptad repeats located at the amino and carboxy-terminal regions which have been proposed to form an intramolecular coiled-coil structure. In a systematic deletion analysis to identify other regions of HSF that may be required to regulate its oligomeric state, we have found that local sequences encompassing the carboxy-terminal end of the DNA binding domain and a broad region of HSF between the heptad repeats also contribute to this regulation. Immunocytochemical analysis of ****mutant**** ****HSF**** proteins revealed a canonical motif required for nuclear localization. HSF proteins lacking the nuclear localization signal remain in the cytoplasm, but these HSFs nonetheless exhibit reversible heat stress-inducible trimerization. The results indicate that the signals that regulate HSF trimerization operate in both the nuclear and cytoplasmic compartments of the cell.

26/7/28 (Item 8 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10563858 BIOSIS NO.: 199699185003
An Hsp70 antisense gene affects the expression of HSP70/HSC70, the regulation of HSF, and the acquisition of thermotolerance in transgenic Arabidopsis thaliana.
AUTHOR: Lee Jeong Hee; Schoeffl Fritz(a)
AUTHOR ADDRESS: (a)Lehrstuhl Allgemeine Genetik, Univ. Tuebingen, Auf der Morgenstelle 28, D-72076 Tuebingen**Germany

JOURNAL: Molecular & General Genetics 252 (1-2):p11-19 1996
ISSN: 0026-8925
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: The genes and proteins of the HSP70 family, are involved in important processes in cells and organelles at normal temperature and after heat stress. Constitutive Hsc 70 and heat-inducible Hsp 70 genes are known in all organisms including plants. The goal of our present investigation was to generate an Hsp70 mutation in Arabidopsis thaliana. In a transgenic approach a heat-inducible antisense Hsp70 gene was constructed, plants were transformed and screened for lack of heat-inducible HSP70 mRNA; two such lines were further investigated. In these plants the Hsp70 gene was not induced by heat shock, and the level of HSC70 RNA was also greatly reduced. This negative antisense effect was specific for genes of the HSP70 family and the induction of mRNAs encoding the small HSP18 class of heat shock protein (HSP) was not affected. The level of HSP70/HSC70 proteins was significantly reduced in transgenic plants, but HSP18 was induced to the same level in different transgenic lines and in untransformed plants. The acquisition of thermotolerance was negatively affected in antisense plants, the survival temperature being 2 degree C below the survival temperature of the wild type and other transgenic lines. Another major effect concerning the regulation of the endogenous heat shock ****transcription**** factor HSF was detected by testing the ability to form heterotrimers between authentic ****HSF**** and ****recombinant**** ****HSF****-GUS (beta-glucuronidase) proteins. The shut-off time, required to turn off HSF activity during recovery from heat stress, was significantly prolonged in antisense plants compared with wild-type and other transgenic lines. Our results imply a dual role of HSP70 in plants, a protective role in thermotolerance and a regulatory effect on HSF activity and hence the autoregulation of the heat shock response.

26/7/29 (Item 9 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10492630 BIOSIS NO.: 199699113775
Role of protein interaction for the activity cycle of ****heat**** ****stress**** ****transcription**** ****factors****.
AUTHOR: Kirchner Christoph; Harmening Uwe; Hoehfeld Ingo; Krishna Priti; Queitsch-Schroedter Christine; Scharf Klaus-Dieter
AUTHOR ADDRESS: Mol. Cell Biology, Biocenter, J.W. Goethe-Univ., Frankfurt/M.**Germany
JOURNAL: European Journal of Cell Biology 69 (SUPPL. 42):p91 1996
CONFERENCE/MEETING: 21st Annual Meeting of the German Society for Cell Biology Hamburg, Germany March 24-28, 1996
ISSN: 0171-9335
RECORD TYPE: Citation
LANGUAGE: English

26/7/30 (Item 10 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10492420 BIOSIS NO.: 199699113565
Functional anatomy of plant ****heat**** ****stress**** ****transcription**** ****factors****.
AUTHOR: Scharf Klaus-Dieter; Treuter Eckardt; Lvck Ruth; Harmening Uwe;

Hoehfeld Ingo; Nover Lutz
AUTHOR ADDRESS: Mol. Cell Biology, Biocenter, J.W. Goethe-Univ., Frankfurt/M.**Germany
JOURNAL: European Journal of Cell Biology 69 (SUPPL. 42):p21 1996
CONFERENCE/MEETING: 21st Annual Meeting of the German Society for Cell Biology Hamburg, Germany March 24-28, 1996
ISSN: 0171-9335
RECORD TYPE: Citation
LANGUAGE: English

26/7/31 (Item 11 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10200039 BIOSIS NO.: 199698654957
Expression of heat shock factor and heat shock protein 70 genes during maize pollen development.
AUTHOR: Gagliardi Dominique; Breton Christian; Chaboud Annie; Vergne Philippe(a); Dumas Christian
AUTHOR ADDRESS: (a)Ecole Normale Superieure Lyon, Reconnaissance Cellularie Amelioration Plantes, UMR CNRS-INRA 993**France
JOURNAL: Plant Molecular Biology 29 (4):p841-856 1995
ISSN: 0167-4412
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: We have analysed the expression of heat shock protein 70 (HSP70) and heat shock factor (****HSF****) ****gene**** during maize pollen development, HSFs being the ****transcriptional**** activators of hsp genes. In order to eliminate the sporophytic tissues of anthers, we have isolated homogeneous cell populations corresponding to five stages of maize pollen development from microspores to mature pollen. We show that in the absence of heat stress, hsp70 genes are highly expressed late-bicellular pollen as compared to other stages. HSP70 transcripts are significantly accumulated in response to a heat shock at the late microspore stage but to a much lower extent than in vegetative tissues. The latest stages of pollen development, i.e. mid-tricellular and mature pollen, do not exhibit heat-induced accumulation of HSP70 transcripts. Therefore, we analysed the expression of ****hsf**** ****genes**** throughout pollen development. We demonstrate that at least three ****hsf**** ****genes**** are expressed in maize and that transcripts corresponding to one ****hsf**** ****gene****, whose expression is independent of temperature in somatic as well as in microgametophytic tissues, are present at similar levels throughout pollen development. In addition, we show that the expression of the two other ****hsf**** ****genes**** is heat-inducible in maize vegetative tissues and is not significantly increased after heat shock at any stage of pollen development. These results indicate that the loss of hsp gene expression at late stages of pollen development is not due to a modification of ****hsf**** ****gene**** expression at the mRNA level and that ****hsf**** ****gene**** expression is differentially regulated in vegetative and microgametophytic tissues.

26/7/32 (Item 12 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10107523 BIOSIS NO.: 199698562441
Stable overexpression of human HSF-1 in murine cells suggests activation

rather than expression of HSF-1 to be the key regulatory step in the heat shock gene expression.
AUTHOR: Mivechi Nahid F(a); Shi Xiao-You; Hahn George M
AUTHOR ADDRESS: (a)Cancer Biol. Res. Lab., Dep. Radiation Oncol., Stanford Univ. Sch. Med., Sanford, CA 94305**USA
JOURNAL: Journal of Cellular Biochemistry 59 (2):p266-280 1995
ISSN: 0730-2312
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: ****Transcription**** of the heat shock genes is regulated by the activation of the heat shock ****transcription**** factor (HSF-1). After heat shock, HSF-1 forms oligomers and binds to the heat shock element (HSE), which consists of several repeats of NGAAN located in the promoter region of the heat shock ****genes****. ****HSF****-1 is then phosphorylated, leading to the enhanced ****transcription**** of the heat shock genes likely by transactivation. We have stably overexpressed the human heat shock ****transcription**** factor-1 (HSF-1) in murine cells to investigate whether the regulation of the expression of the heat shock genes may partly reside at the level of HSF-1 expression. Human HSF-1 cDNA was cloned into a retroviral vector (pvhhsf-1) and was overexpressed in a murine fibroblast cell line. The overexpressed human HSF-1 is found in both the cytoplasm and nucleus of control cells but is translocated into the nucleus upon heat shock. Electrophoretic mobility shift analysis suggests that the human HSF-1 has constitutive DNA binding ability and its DNA binding ability is increased upon heat shock. Cross-linking experiments indicate that the overexpressed human HSF-1 is mainly a monomer under control conditions and forms oligomers upon heat shock. Immunoblotting shows that the human HSF-1 is phosphorylated upon heat shock and its apparent molecular weight is shifted up by at least 10 kDa. In spite of both the DNA binding ability and phosphorylation, the overexpression of human HSF-1 does not increase the ****transcription**** of murine HSP-70 mRNA or increase the synthesis of other HSPs after heat shock beyond that observed in control untransfected cells. An exception is the enhanced synthesis of a 47-50 kDa protein after heat shock and an apparent lack of induction of one HSP-70 kDa species when the protein pattern is analyzed by isoelectric focusing. Interestingly, cells overexpressing human HSF-1 show a 4-fold increase in the basal expression of luciferase when the plasmids containing the human HSP-70 promoter ligated to the luciferase reporter gene are transiently expressed in these cells. Murine cells overexpressing human HSF-1 are more resistant to the cytotoxic effects of heat when compared to the control untransfected cells, but the kinetics of thermotolerance development and decay is similar between HSF-1 transfected and untransfected cells. In conclusion, human HSF-1 protein in murine fibroblasts is modified in a similar fashion as the endogenous mouse HSF-1 after heat shock. However, the overexpression of HSF-1 does not result in overproduction of heat shock proteins after heat shock, perhaps because these cells contain abundant amounts of endogenous HSF-1.

26/7/33 (Item 13 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

10060743 BIOSIS NO.: 199598515661
The DNA-binding properties of two heat shock factors, HSF1 and HSF3, are induced in the avian erythroblast cell line HD6.
AUTHOR: Nakai Akira(a); Kawazoe Yoshinori; Tanabe Masako; Nagata Kazuhiro; Morimoto Richard I
AUTHOR ADDRESS: (a)Dep. Cell Biol., Chest Disease Res. Inst., Kyoto Univ.,

Sakyo-ku, Kyoto 606**Japan
JOURNAL: Molecular and Cellular Biology 15 (10):p5268-5278 1995
ISSN: 0270-7306
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Avian cells express three heat shock ****transcription**** factor (****HSF****) ****genes**** corresponding to a novel factor, HSF3, and homologs of mouse and human HSF1 and HSF2. Analysis of the biochemical and cell biological properties of these HSFs reveals that HSF3 has properties in common with both HSF1 and HSF2 and yet has features which are distinct from both. HSF3 is constitutively expressed in the erythroblast cell line HD6, the lymphoblast cell line MSB, and embryo fibroblasts, and yet its DNA-binding activity is induced only upon exposure of HD6 cells to heat shock. Acquisition of HSF3 DNA-binding activity in HD6 cells is accompanied by oligomerization from a non-DNA-binding dimer to a DNA-binding trimer, whereas the effect of heat shock on HSF1 is oligomerization of an inert monomer to a DNA-binding trimer. Induction of HSF3 DNA-binding activity is delayed compared with that of HSF1. As occurs for HSF1, heat shock leads to the translocation of HSF3 to the nucleus. HSF3 exhibits the properties of a ****transcriptional**** activator, as judged from the stimulatory activity of transiently overexpressed HSF3 measured by using a heat shock element-containing reporter construct and as independently assayed by the activity of a chimeric GAL4-HSF3 protein on a GAL4 reporter construct. These results reveal that HSF3 is negatively regulated in avian cells and acquires DNA-binding activity in certain cells upon heat shock.

26/7/34 (Item 14 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

09969488 BIOSIS NO.: 199598424406
Dual regulation of the Drosophila hsp26 promoter in vitro.
AUTHOR: Sandaltzopoulos Raphael; Mitchelmore Catherine; Bonte Edgar; Wall Gayl; Becker Peter B(a)
AUTHOR ADDRESS: (a)Gene Expression Programme, European Mol. Biol. Lab., Meyerhofstrasse 1, 69117 Heidelberg**Germany
JOURNAL: Nucleic Acids Research 23 (13):p2479-2487 1995
ISSN: 0305-1048
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Efficient heat shock induction of Drosophila hsp26 gene ****transcription**** in vivo requires binding sites for heat shock factor (HSF) and GAGA factor (GAF) close to the TATA box (proximal elements) as well as 350 bp upstream of the start site of ****transcription**** (distal elements). We have evaluated the contribution of hsp26 promoter sequences to ****transcriptional**** activity in extracts from either heat shocked or unstressed fly embryos. Efficient ****transcription**** in either extract was governed by distinct regulatory principles. ****Transcription**** in extracts from unstressed embryos relied solely on GAGA elements which efficiently counteracted repression by abundant non-specific DNA-binding proteins. ****Transcription**** in extracts from heat shocked embryos depended only a little on GAGA elements, relying mainly on functional HSEs. Constitutively active ****recombinant**** ****HSF**** or native factor in an extract from heat shocked embryos was able to truly activate ****transcription**** essentially via proximal HSEs, but not when bound

to distal sites. These two modes of regulation in vitro may correspond to the two functional states of the promoter before and after heat shock in vivo.

26/7/35 (Item 15 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

09952774 BIOSIS NO.: 199598407692
Cooperative Binding of Heat Shock ****Transcription**** Factor to the Hsp70 Promoter in Vivo and in Vitro.
AUTHOR: Amin Jahanshah; Fernandez Mary; Ananthan Jayakumar; Lis John T; Voellmy Richard(a)
AUTHOR ADDRESS: (a)Dep. Biochem. Mol. Biol., Univ. Miami Sch. Med., PO Box 016129, Miami, FL 33101-1019**USA
JOURNAL: Journal of Biological Chemistry 269 (7):p4804-4811 1994
ISSN: 0021-9258
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: The minimal promoter of the Drosophila hsp70 gene contains a TATA box and two nonidentical HSE sequences, HSEI and HSEII, that synergistically activate the promoter. We have examined stereospecific alignment and spatial constraints in this promoter. Similar to deletion of HSEII, insertion in the spacer between the HSEs of 1 to 5 or 11 to 14 nucleotides (nt) reduced promoter activity to about 10%. In contrast, HSEII was capable of contributing to promoter activity when the spacer was either shortened by 2 or 4 nt or extended by 6 to 10 or 16 or 18 nt. Hence, half of the possible helical arrangements of HSEs are compatible, whereas the other half are essentially incompatible with efficient promoter function. HSEII was ineffective when its distance to HSEI was increased by more than 18 nt. In vitro, HSEII is a weak and HSEI a strong binding site for heat shock ****transcription**** factor HSF, and HSF binds to HSEII cooperatively. To find out whether the above periodicity reflects cooperative binding of HSF in vivo or represents the need of stereoalignment for synergistic activation of ****transcription****, the weak HSF binding site HSEII was replaced with the strong binding site HSEI. This substitution greatly attenuated promoter periodicity, suggesting that the periodic effects are caused by cooperative binding of HSF to HSEII, and that stereoalignment of HSEs is not required for ****transcription**** activation. In agreement, in vitro assays using spacer mutants revealed cooperative binding of purified, ****recombinant**** ****HSF**** to HSEII with a similar periodicity as observed in vivo. Changing the distance between TATA and the HSEs did not produce promoter periodicity, indicating that stereoalignment of these elements is not important.

26/7/36 (Item 16 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

09871946 BIOSIS NO.: 199598326864
Estrogen dependent expression of heat shock ****transcription**** factor: Implications for uterine synthesis of heat shock proteins.
AUTHOR: Yang Xinli; Dale Emily C; Diaz Jaime; Shyamala G(a)
AUTHOR ADDRESS: (a)Dep. Cell Molecualr Biol., Lawrence Berkeley Lab., Univ. California, Berkeley, CA 94720**USA
JOURNAL: Journal of Steroid Biochemistry and Molecular Biology 52 (5):p 415-419 1995

ISSN: 0960-0760
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: ****Transcriptional**** induction of heat shock protein genes is generally mediated by binding of heat shock ****transcription**** factor(s) to the heat shock element present in the promoters of heat shock genes. Although the steady-state levels of heat shock factor mRNAs vary among different tissues, at present virtually nothing is known regarding the cellular signals responsible for their synthesis and hence the observed variations. In this report we demonstrate that the heat shock ****transcription**** factor (HSTF or HSF) is under positive regulation by estrogen. The effect of estrogen was observed with both types of heat shock factors (HSF-1 and HSF-2) and occurred at both the mRNA and protein level. Immunolocalization studies emphasized the potential biological importance of these observations whereby the increase in uterine HSF-1 and HSF-2 due to estrogen was found to be associated with the endometrium, the primary tissue component which is targeted for estrogen action. This is the first demonstration of a cellular factor which can regulate ****HSF****-1 and ****HSF****-2 ****gene**** expression. The implications of these findings to uterine heat shock protein gene expression are discussed.

26/7/37 (Item 17 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

09725758 BIOSIS NO.: 199598180676
In vitro activation of purified human heat shock factor by heat.
AUTHOR: Larson Jeffrey S; Schuetz Thomas J; Kingston Robert E(a)
AUTHOR ADDRESS: (a)Mass. General Hosp., Wellman 10, Blossman Street, Boston, MA 02114**USA
JOURNAL: Biochemistry 34 (6):p1902-1911 1995
ISSN: 0006-2960
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: A major regulatory step in the heat-induced ****transcription**** of heat shock protein (hsp) genes in eukaryotes is the activation of heat shock factor (HSF). In metazoans and Schizosaccharomyces pombe, HSF is present in unstressed cells but is unable to bind to its target DNA sequence element, the heat shock element (HSE). Heat induction of the DNA binding activity of HSF is a critical component required for activation of heat shock ****genes****. Inactive ****HSF**** in extracts of non-heat shocked human cells can be heated in vitro to activate HSF, suggesting the factors required to sense temperature and activate HSF are soluble factors (Larson, J. S., Schuetz, T. J., & Kingston, R. E. (1988) Nature 335, 372-375). We utilized the ability to purify human HSF in the active form to characterize further the in vitro activation of HSF. Here we have developed a procedure to deactivate the DNA binding ability of HSF. When purified and deactivated HSF is heated, the DNA binding ability of HSF is activated. This activation occurs most efficiently at 43 degree C (heat shock temperature), but, in contrast to activation in the crude system, some activation of HSF is observed at 37 degree C (non-heat shock temperature). We show that purified and deactivated HSF is similar to natural inactive HSF in both size and shape. Thus, the monomer to trimer transition that activates HSF can occur in a temperature-dependent fashion in the absence of other proteins. It is possible that these biochemical properties of HSF contribute to the abilitv of HSF to respond

to heat in vivo.

26/7/38 (Item 18 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

09707853 BIOSIS NO.: 199598162771
Ectopic expression of ****chimaeric**** ****HSF**** causes constitutive activation of HSP synthesis in transgenic Arabidopsis plants.
AUTHOR: Schoeffl Fritz; Lee Jeong H; Huebel Anja
AUTHOR ADDRESS: Lehrstuhl Allgemeine Genetik, Univ. Tuebingen, Auf Morgenstelle 28, D-72074 Tuebingen**Germany
JOURNAL: Journal of Cellular Biochemistry Supplement 0 (19B):p204 1995
CONFERENCE/MEETING: Keystone Symposium on Heat Shock (Stress) Proteins in Biology and Medicine Santa Fe, New Mexico, USA February 27-March 5, 1995
ISSN: 0733-1959
RECORD TYPE: Citation
LANGUAGE: English

26/7/39 (Item 19 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

09356850 BIOSIS NO.: 199497365220
Analysis of HSF-1 phosphorylation in A549 cells treated with a variety of stresses.
AUTHOR: Mivechi N F(a); Koong A C; Giaccia A J; Hahn G M
AUTHOR ADDRESS: (a)Dep. Radiation Oncol., Cancer Biol. Research Lab., Stanford, CA 94305**USA
JOURNAL: International Journal of Hyperthermia 10 (3):p371-379 1994
ISSN: 0265-6736
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: In the absence of stress. heat shock transcript ion factor-1 (HSF-1) exists as a monomer. After the treatment of cells with variety of stresses, HSF-1 forms a trimer and binds to the heat shock element (HSE), a motif consisting of three consecutive NGAAN sequences located in an inverted orientation upstream of the heat shock ****genes****. ****HSF****-1 is then phosphorylated causing transactivation of heat shock mRNAs. Treatment of cells with some of the stresses has been shown to increase HSF binding to HSE without detectably increasing the synthesis of heat shock mRNAs. Here we used antibody specific to HSF-1 to detect its phosphorylation status following exposure of A549, a human lung carcinoma cell line to a variety of stresses in order to correlate HSF-1 phosphorylation with its transactivation ability. Our studies show that HSF-1 is phosphorylated following heat shock (43 degree C for 1 h) hypoxia (5 h exposure to 0.02% oxygen), 8% ethanol (1 h exposure at 37 degree C). or 200 mu-M sodium arsenite (1 h exposure at 37 degree C). All such stresses have previously been shown to increase the synthesis of heat shock proteins (hsps). However, there are no detectable increases in HSF-1 phosphorylation after the treatment of cells with X-irradiation (2-8 Gy) or 100 mu-M canavanine, an amino acid analogue (1 h exposure at 37 degree C). Treatment of cells with X-irradiation increases HSF binding to HSE without increasing the synthesis of hsps, while treatment of cells with canavanine has been shown to increase the synthesis of hsps. These results suggest that with the exception of amino acid analogues, all stresses which cause an increase in HSF-1 phosphorylation also enhance the synthesis of hsps, We also attempted to inhibit phosphorylation of

HSF-1 with protein kinase inhibitors. Neither herbimycin A, a general protein tyrosine kinase inhibitor, nor staurosporine, a ser/thr protein kinase inhibitor interfered with the phosphorylation of HSF-1 in A549 cells.

26/7/40 (Item 20 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

08819107 BIOSIS NO.: 199395108458
Activation of heat shock gene ****transcription**** by heat shock factor 1 involves oligomerization acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress.
AUTHOR: Sarge Kevin D; Murphy Shawn P; Morimoto Richard I(a)
AUTHOR ADDRESS: (a)Dep. Biochem. Mol. Biol. and Cell Biol., Northwestern Univ., Evanston, IL 60208**USA
JOURNAL: Molecular and Cellular Biology 13 (3):p1392-1407 1993
ISSN: 0270-7306
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: The existence of multiple heat shock factor (****HSF****) ****genes**** in higher eukaryotes has prompted questions regarding the functions of these HSF family members, especially with respect to the stress response. To address these questions, we have used polyclonal antisera raised against mouse HSF1 and HSF2 to examine the biochemical, physical, and functional properties of these two factors in unstressed and heat-shocked mouse and human cells. We have identified HSF1 as the mediator of stress-induced heat shock gene ****transcription****. HSF1 displays stress-induced DNA-binding activity, oligomerization, and nuclear localization, while HSF2 does not. Also, HSF1 undergoes phosphorylation in cells exposed to heat or cadmium sulfate but not in cells treated with the amino acid analog L-azetidine-2-carboxylic acid, indicating that phosphorylation of HSF1 is not essential for its activation. Interestingly, HSF1 and HSF2 overexpressed in transfected 3T3 cells both display constitutive DNA-binding activity, oligomerization, and ****transcriptional**** activity. These results demonstrate that HSF1 can be activated in the absence of physiological stress and also provide support for a model of regulation of HSF1 and HSF2 activity by a titratable negative regulatory factor.

26/7/41 (Item 21 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

08562951 BIOSIS NO.: 199344112951
Cloning, expression and functional analysis of three plant ****heat**** ****stress**** ****transcription**** ****factor**** genes.
AUTHOR: Scharf K D; Treuter E; Rose S; Nover L
AUTHOR ADDRESS: Inst. Plant Biol., 0-4050 Halle**Germany
JOURNAL: Journal of Cellular Biochemistry Supplement 0 (17 PART A):p90 1993
CONFERENCE/MEETING: Keystone Symposium on Transcription: Factors, Regulation and Differentiation Keystone, Colorado, USA January 17-24, 1993
ISSN: 0733-1959
RECORD TYPE: Citation
LANGUAGE: English

26/7/42 (Item 22 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2000 BIOSIS. All rts. reserv.

07453715 BIOSIS NO.: 000091049934
MOLECULAR CLONING AND EXPRESSION OF A HEXAMERIC DROSOPHILA HEAT SHOCK FACTOR SUBJECT TO NEGATIVE REGULATION
AUTHOR: CLOS J; WESTWOOD J T; BECKER P B; WILSON S; LAMBERT K; WU C
AUTHOR ADDRESS: LAB. BIOCHEM., NATIONAL CANCER INST., NATIONAL INST. HEALTH, BETHESDA, MD. 20892.
JOURNAL: CELL 63 (5): 1990. 1085-1098.
FULL JOURNAL NAME: Cell
CODEN: CELLB
RECORD TYPE: Abstract
LANGUAGE: ENGLISH

ABSTRACT: We report the cloning of the ****transcriptional**** activator of heat shock ****genes****, ****HSF****, from Drosophila. The predicted sequence of Drosophila HSF protein is surprisingly divergent from that of yeast HSF, except in regions important for DNA binding and oligomerization. A segment of the DNA binding domain of HSF bears an intriguing similarity to the putative DNA recognition helix of bacterial sigma factors, while the oligomerization domain contains an unusual arrangement of conserved hydrophobic heptad repeats. Drosophila HSF produced in Escherichia coli under nonshock conditions forms a hexamer that binds specifically to DNA with high affinity and activates ****transcription**** from a heat shock promoter in vitro. In contrast, when HSF is expressed in Xenopus oocytes, maximal DNA binding affinity is observed only after heat shock induction. These results suggest that Drosophila HSF has an intrinsic affinity for DNA, which is suppressed under nonshock conditions in vivo.

26/7/43 (Item 1 from file: 73)
DIALOG(R)File 73:EMBASE
(c) 2000 Elsevier Science B.V. All rts. reserv.

04632644 EMBASE No: 1991126687
Erratum: Three tomato genes code for ****heat**** ****stress**** ****transcription**** ****factors**** with a region of remarkable homology to the DNA-binding domain of the yeast HSF (Klaus-Dieter Scharf, Sonja Rose, Wolfgang Zott, Fritz Schoffl and Lutz Nover) (The EMBO Journal, 9, 4495-4501, 1990)
EMBO Journal (EMBO J.) (United Kingdom) 1991, 10/4 (1026)
CODEN: EMJOD ISSN: 0261-4189
DOCUMENT TYPE: Journal; Erratum
LANGUAGE: ENGLISH

THIS PAGE BLANK (USPTO)